Se 4

DEVELOPMENT OF A SELF-ASSESSMENT EXAMINATION FOR PHYSICIAN ASSISTANTS

NEWARK PUBLIC LIBRARY

SEP 26 1980

U. S. GOVT. PUBS.
DEPOSITORY 376

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES • Public Health Service • Health Resources Administration

HE 20.6002:

DEVELOPMENT OF A SELF-ASSESSMENT EXAMINATION FOR PHYSICIAN ASSISTANTS

Donald W. Fisher, Ph.D.
Executive Director
American Academy of Physician Assistants

Ayres D'Costa, Ph.D.
Associate Professor
The Ohio State University

American Academy of Physician Assistants
Arlington, Virginia

August 1979

Contract No. 231-76-0053

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
Public Health Service
Health Resources Administration
Bureau of Health Professions

DHHS Publication No. (HRA) 80-38

FOREWORD

Developing continuing education materials to assist health practitioners maintain basic competencies and acquire new competencies is a challenge for the health professions. Central to this theme is a recognition of the importance of identifying the continuing education needs of a profession.

In recognition of this problem, the Health Professions Educational Assistance Act of 1976 (P.L. 94-484) authorizes the establishment of continuing education programs for practicing health personnel. This report, prepared by the American Academy of Physician Assistants of Arlington, Virginia, addresses the issue of developing a self-assessment system designed to identify continuing education needs of physician assistants.

In this report a unique self-assessment system incorporating practitioner's practice pattern, self-estimates of need, and knowledge level in basic competency domains essential to practice is described. The resulting material, we feel, will aid not only physician assistants, but all health professionals interested in continuing education. We are pleased to have been able to assist in this effort.

David B Hoover

David B. Hoover
Acting Director
Division of Associated Health Professions
Bureau of Health Professions
Health Resources Administration

July 1980

TABLE OF CONTENTS

FOREWORD.... iii

ACKNOWLEDGMENTS.... x

PREFACE.... xii

ABSTRACT.... xxii

I. INTRODUCTION.... 1

- A. Background to Project.... 1
- B. Purpose of Report.... 1
- C. Rationale of Project.... 3

II. PROJECT PROCEDURES.... 7

- A. Strategy.... 7
- B. Work Plan.... 8
- C. An Overview of the Actual Work Schedule.... 10
- D. List of Products Developed.... 16
- E. Project Problems Encountered.... 17

III. TECHNICAL ISSUES.... 20

- A. Objectives of Self-Assessment Examination.... 20
- B. Technical Rationale of Test.... 24
- C. Test Specifications.... 25
- D. Scales Definition.... 31
- E. Item Generation.... 38
- F. Test Item Revision.... 47
- G. Test Development Statistics.... 53
- H. Test Interpretation.... 54
- I. A Review of Technical Limitations and Deficiencies.... 64

IV. USING THE SELF-ASSESSMENT EXAMINATION TO DEVELOP A PILOT CME SYSTEM FOR PAs.... 66

- A. Rationale.... 66
- B. Methodology Proposed.... 68
- C. Time Schedule.... 71

V. FINDINGS AND CONCLUSIONS.... 72

REFERENCES.... 77

READEX FICHE

LIST OF TABLES

I. Revised Schedule for Test Development.................... 11

II. Test Specifications Matrix Cells and Marginals........... 28

III. Description of the 17 PA Role Area Scales................ 32

IV. Codes/Levels for Classification Variables................ 35

V. How to Classify Test Items by Skills Levels.............. 39

VI. Assignment of Test Section 1 Items to Specifications Matrix... 40

VII. Guide for Generating Test Item Scenarios................. 41

VIII. Sample Scenario.. 42

IX. Sample Test Item Based on Scenario....................... 43

X. Evaluative Data... 76

LIST OF FIGURES

1. The Four Types of Expectations of Health Professionals.......... 23

2. Implementing the Test Specifications Matrix Using TS Committee Marginals.......... 30

3. Map of the 17 Role Areas X Body System Scales.......... 33

4. Map of the 28 Body System X Medical Intervention Scales.......... 34

5. Sample Test Item From Item Bank.......... 48

6. Test Item Documentation Card Listing.......... 49

LIST OF EXHIBITS

A. Unprioritized List of Program and Research Objectives.................................... 79

B. The Eleven Major Responsibilities of the PA Role Delineation.......................... 80

C. Frequency Distribution of Total Scores.................. 83

D. Mean, Standard Deviation, and Minimum Competency Score for the 17 Role Scales......................... 84

E. Mean, Standard Deviation, and Minimum Competency Score for the 28 System Scales....................... 85

F. Distributions of Item Difficulty and Discrimination Indices for Total Test (Second Try-Out).................................. 86

G. Reliability, Standard Error of Measurement (SEM), and Coefficient of Acceptance for the 17 PA Role Scales.. 87

H. Reliability, Standard Error of Measurement (SEM), and Coefficient of Acceptance for the 28 Body System Scales................................. 88

I. Assignment of Second Try-Out Test Items to Test Specifications Matrix................................ 89

J. Individual Report.................................. 90

K. Interpretive Leaflet................................ 94

ABBREVIATIONS

AAPA	American Academy of Physician Assistants
APAP	Association of Physician Assistant Programs
CME	Continuing Medical Education
HRA	Health Resources Association
IWC	Item Writers Committee
MCV	Medical College of Virginia Study (Marsland et al., 1976)
NCHSR	National Center for Health Services Research
PA	Physician Assistant
PAs	plural of PA
PMP	Patient Management Problem
TSC	Test Specifications Committee

ACKNOWLEDGMENTS

I wish to thank the project staff, consultants, and committee members (listed in the Preface) who gave so much of their time and experience. We owe a large debt of gratitude to Louis Quatrano, Ph.D., Margaret Wilson, Ph.D., and Robert Conant, Ph.D., for their contributions, advice and leadership throughout the project.

A special thanks, however, goes to those persons who worked as staff on contract activities for a limited period of time, particularly Judy A. Light for her creativity in research design, guidance, perservance, and dedication to the project and the physician assistant profession during her tenure as project director (1976-1978). Also, professional staff support was provided by Ted Kastelic, Ph.D., and secretarial and administrative staff who worked efficiently and effectively on a variety of contract activities were Laura Friedman and Susan Herre.

The staff for the contract would also like to thank others working in the National Office who took time out of busy schedules to review reports, coordinate activities, and offer suggestions. Such cooperation is appreciated.

A very special thanks goes to members of the physician assistant profession. The willingness of physician assistants to spend long hours completing forms, taking tests, and serving on committees is testimony to the enthusiasm and commitment of these individuals to their profession and its objectives.

Donald W. Fisher, Ph.D.

ACKNOWLEDGMENT

This report was prepared for the American Academy of Physician Assistants by the Project Consultant, Ayres D'Costa, Ph.D., Associate Professor, Health Professions Education, The Ohio State University, Columbus, Ohio.

PREFACE

Participants in all parts of the study described in the five volumes of this report are listed below.

The project staff and their areas of responsibility were:

Name	Responsibility
Donald W. Fisher, Ph.D. Executive Director, AAPA/APAP	Project Director
Mary Jane Crain Research Associate	Assessment of the applicability of the University of Wisconsin's Individual Physician Profile (IPP) program for physician assistants
	Maintenance of a roster of CME programs for physician assistants
	Design of a system of CME program accreditation
Jane Faulman, Ph.D. Research Associate	Verification of the role delineation for the entry level generalist position
	Physician assistant position classification
	Development of a self-assessment tool

The Project Officer for the study, from the Division of Associated Health Professions, was Louis A. Quatrano, Ph.D.

Secretarial and administrative support was provided by:

Veronica Marshall
Karen Hummer
Linda Geary

The project staff consulted a measurement and evaluation specialist, Dr. Richard C. Cox of Pittsburgh, Pennyslvania, to assist with the IPP and system of continuing medical education (CME) program accreditation portions of this study. Specifically, Dr. Cox designed a checklist for IPP participants and helped with the analysis and presentation of all IPP checklist resullts. He also assisted the staff in the development of checklists distributed to physician assistants in attendance at selected CME programs throughout the country. The data from these checklists was used in the development of a system of accreditation of physician assistant oriented CME programs.

The development of a self-assessment tool was placed under the direction of a consultant to the project who is a specialist in test development. This consultant was Ayres D'Costa, Ph.D., Associate Professor of Health Professions Education at The Ohio State University in Columbus, Ohio. Under Dr. D'Costa's guidance, a self-assessment examination for physician assistants was developed. Dr. D'Costa planned and conducted all meetings at which the test specifications for the examination were delineated and test items were prepared and revised. He was responsible for all computer output necessary to the project. He designed an individualized, computer-generated test report which includes respondents' scale scores both numerically and graphically.

Two other consultants worked with Dr. D'Costa in the training of the item writers and in developing training materials and test items for the self-assessment examination. These consultants were Roger Trent, Ph.D., and Cherry Turner, Ph.D. Dr. Trent is responsibile for testing programs with the Ohio Department of Education, and Dr. Turner is affiliated with the Ohio State University College of Medicine.

In addition to the help of consultants, the project staff also benefited from the special expertise and insight of members of the Evaluation, Working, and Advisory committees. Each of the committees had a specific role to play in the completion of this study.

The Evaluation Committee worked primarily on the assessment of the applicability of IPP for physician assistants, the design of a system of CME accreditation, and the maintenance of a roster of CME programs. Members of this committee reviewed the data collected about the Individual Physician Profile program, suggested other information to be obtained, and made recommendations regarding the program's applicability for physician assistants. The Evaluation Committee had a major role in the development of instruments, the review of data, and the making of recommendations regarding continuing medical educatin options and accreditation systems. Also, this committee reviewed the roster format for CME programs.

Members of the Evaluation Committee included educational specialists competent in criterion-referenced measurement, design of instructional materials, evaluation methodology, and clinical simulation. The Academy's Professional and Continuing Education Committee had two representatives serving on the Evaluation Committee. The ten members of the Evaluation Committee were:

Philip G. Bashook, Ed.D
Chicago, Illinois

Robert J. Blakely
Chicago, Illinois

Sarah M. Dinham, Ph.D.
Tucson, Arizona

Stephen C. Gladhart, Ph.D.
Wichita, Kansas

Thomas R. Godkins, P.A.
Oklahoma City, Oklahoma

Jan L. Hagen, M.S.W.
Baltimore, Maryland

Paul F. Moson, P.A.-C.
Loretto, Pennsylvania

Robert R. Moutrie, Ph.D.
Newark, New Jersey

John E. Ott, M.D.
Washington, D.C.

Paul S. Toth, P.A.-C.
Durham, North Carolina

The seven-member Working Committee worked closely with the project staff in developing the Role Delineation for the Physician Assistant. This document was produced via the accomplishment of two tasks: verification of an earlier role delineation (included in the Curriculum Resource Document project) and determination of a position classification for the physician assistant profession. Members of this committee included practicing physician assistants, physicians (in private practice and hospital settings) who employ physician assistants, faculty of physician assistant training programs and one representative from the Curriculum Resource Document project. The members of the Working Committee were:

Mack Bonner, Jr., M.D.
New York, New York

Trudy Jo Companiotte, P.A.-C.
Nashville, Tennessee

William E. G. de Alva, M.D.
Denver, Colorado

Carl E. Fasser, P.A.-C.
Houston, Texas

Stephen L. Joyner, P.A.-C.
Ayden, North Carolina

Allan B. Kunkel, M.D.
Cleveland, Ohio

Daniel O. Myhre, P.A.-C.
Spokane, Washington

Representatives from major medical organizations with a significant interest in the physician assistant profession served on an Advisory Committee to review materials and provide input to the staff and the other two committees for all phases of the study. Representatives from the two other committees for the contract served as liaison members on the Advisory Committee. The members of this committee reviewed and provided advice on data, interim reports, and conclusions and recommendations about the role delineation for the physician assistant, the Individual Physician Profile program, the system of the CME program accreditation, and the roster of CME programs for physician assistants. The members of this Advisory committee were:

Leo S. Bell, M.D., F.A.A.P., F.A.P.H.A.
San Mateo, California
American Academy of Pediatrics

Pearl H. Dunkley, R.N., Ed.D.
Kansas City, Missouri
American Nurses' Association

Carl E. Fasser, P.A.-C.
Houston, Texas
Representative of the Working Committee

Dan P. Fox, P.A.-C.
Oklahoma City, Oklahoma
American Academy of Physician Assistants

Thomas R. Godkins, P.A.
Oklahoma City, Oklahoma
Representative of the Evaluation Committee

Rolf M. Gunnar, M.D., F.A.C.P.
Maywood, Illinois
American Medical Association

J. Rhodes Haverty, M.D.
Atlanta, Georgia
National Commission on Certification of Physician's Assistants

Frances L. Horvath, M.D.
St. Louis, Missouri
Association of Physician Assistant Programs

Joseph A. Intile, Jr., M.D., F.A.C.P.
Oregon City, Oregon
American Society of Internal Medicine

Robert E. Jewett, M.D.
Dayton, Ohio
Association of American Medical Colleges

Raymond H. Murray, M.D., F.A.C.P.
East Lansing, Michigan
American College of Physicians

Dan A. Nye, M.D.
Kearney, Nebraska
Federation of State Medical Boards
of the United States

Frederic L. Schoen, M.D.
Indianapolis, Indiana
American Academy of Family Physicians

Daniel R. Thomas*
Chicago, Illinois
American Hospital Association

Harold Zintel, M.D., F.A.C.S.
Chicago, Illinois
American College of Surgeons

Three groups of physician assistant practitioners and educators contributed to the development of the Self-Assessment Examination for Physician Assistants. The Test Specifications Committee provided input for the test specifications matrix, for item revision, and for future research. The six members of this Committee were:

Carl E. Fasser, PA-C
Houston, TX 77031

David Glazer, M.A.
Atlanta, GA 30327

Allan B. Kunkel, M.D.
Painesville, OH 44077

Laurie Lipsig, PA-C
Chicago, IL 60610

Thomas E. Piemme, M.D.
Washington, DC 20037

---

*In May 1979, Thomas Atchison, Ed.D., replaced Daniel Thomas at the American Hospital Association.

Judith B. Willis, M.A., PA-C
Kalamazoo, MI 49007

A committee of 24 physician assistants met in two workshops to develop and revise test items. The large majority of the items on the exam were produced by this group. Workshop Item Writers were:

Donald A. Abrams, PA
Jamaica Plain, MA 02130

Randall C. Bennett, PA-C
Trenton, NJ 32693

Scott Chavez, PA-C
Las Vegas, NE 89119

Robert Christie, PA-C
Dayton, OH 45449

Linda K. Davies, PA-C
Arlington, VA 22206

Dale B. Davis, PA-C
Springfield, MO 65807

Max R. Dawkins, PA-C
Greensburg, IN 47240

Robert L. France, PA-C
Sidney, MT 95270

Edward B. Friedmann, PA-C
Mason City, IA 50401

George F. Hillegas III, PA-C
Baltimore, MD 21226

Norman Holton, PA-C
St. Clair Shores, MI 48081

Charles E. Horan, PA-C
Phoenix, AZ 85032

Paul Lombardo, PA-C
W. Babylon, NY 11704

John McCarty, PA
Marshfield, WI 54449

Noel H. McFarlane, PA-C
Silver Spring, MD 20906

Dennis W. O'Dell, PA-C
Wailuku Maui, HA 96793

Leonard T. O'Neill, PA-C
Omaha, NE 68134

Kenneth Ryther, PA-C
Delta Junction, AK 99737

Michael Sheldon, PA-C
Portland, ME 04103

Valerie Staples, PA-C
Hillsborough, NC 27272

Valgene Valgora, PA-C
Omaha, NE 68105

Joseph Varano, PA-C
Philadelphia, PA 19103

Cecil Walker, PA-C
Carson, CA 90746

L. Timothy Whitmore, PA-C
Richmond, VA 23235

Thirty PAs were asked to be Field Item Writers. Materials on writing test items were mailed to them, and they were requested to write items and forward them to the national office. The PAs asked to be Field Item Writers were:

Timothy Bauer, PA-C
Tomah, WI 54660

Walker Boone, PA-C
Asheville, NC 28804

Paul Cephus, PA-C
Houston, TX 77089

Michelle Combs, PA-C
Lexington, KY 40504

Wayne C. Cure
Ravena, NY 12143

Laura Davis-Montgomery, PA-C
Advance, NC 27006

Marc Dicker, PA-C
Wichita, KS 67226

David Fraser, PA-C
Denton, TX 76201

Gale Harkness, PA-C
Winston-Salem, NC 27105

Marcia Hawkins, PA-C
Virginia Beach, VA 23456

William H. Hoge, PA-C
Albuquerque, NM 87112

William A. Holefelder, PA-C
Wichita, TX 76306

Anthony Ilardi, PA-C
Inwood, WV 25428

Glenna R. Jones, PA-C
Lexington, KY 40503

Raymond J. Krystyniak, PA-C
Chicago, IL 60657

Michael Kuns, PA-C
Oklahoma City, OK 73120

Robert Lapham, PA-C
Portland, ME 04102

Kathleen Lester, PA-C
Seattle, WA 98105

Kederick A. Meyers, PA-C
Shelbyville, IN 46176

Leonard Milcowitz, PA-C
New Port Richey, FL 33552

Sandra E. Mitchell, PA-C
Eugene, OR 97401

Thomas N. Nalley, PA-C
Doraville, GA 30360

Donald Price, PA-C
Florissant, MO 63031

Joseph E. Ricketts, Jr., PA-C
Centerville, OH 45459

Richard H. Salladin, PA-C
Berea, OH 44017

Kenneth L. Smith, PA-C
Carrollton, GA 30117

Sandra Sommers, PA-C
Towanda, KS 67144

Gayle Spears, PA-C
Denver, CO 80203

Steven Tiger, PA-C
Jamaica, NY 11432

ABSTRACT

The purpose of this project was to develop a criterion-referenced self-assessment examination for physician assistants (PAs), using the Role Delineation as the basis, from which appropriate continuing education packages could be developed. The test development effort was undertaken with the help of Working Committees consisting of PAs and PA educators. A six-hour examination consisting of 315 items has been constructed using two try-outs.

The domain of the examination is the competency skills and knowledge expected of an entry-level generalist PA. The domain has been defined in terms of two sets of scales: 17 Role Scales and 28 Body System Scales. The interpretation of scores is based upon minimum competency scores decided upon by expert judgement using the Nedelsky Technique.

Two innovative approaches were used in the implementation of this project. One involved the use of critical incidents in the generation of test items. The other involved the use of a three-factor conceptual model for continuing medical education (CME) using self-assessment examinations. It is the thesis of this model that CME must be based upon a combined analysis of practice (P) requirements, individual felt needs (N), and deficits identified by examiniation (E) scores. A four-page computer generated report was developed and returned along with an Interpretive Leaflet as feedback to each PA who participated in the Try-Out Examination of June 1979.

I. INTRODUCTION

A. *Background to Project*

The Self-Assessment project of the American Academy of Physician Assistants (AAPA) lies at the very core of its mission "to facilitate the recognition of the physician assistant as a professional dedicated to the delivery of quality care" (AAPA, 1978). Quality looms as a major concern of this new profession. The National Center for Health Services Research (NCHSR, 1978) cited a 1976 estimate indicating that 1200 physician assistants and nurse practitioners had been trained as a result of federal support since 1965. This 1978 NCHSR Report on nurse practitioners and physician assistants focuses on medical care utilization issues, particularly those emanating from current insurance reimbursement restrictions. The NCHSR Report recommended an interim 100 percent reimbursement based upon the principle that (reimbursement) rates should be related to the service performed and not who performs the service.

Quality medical care is based on the competence of the provider, but it also recognizes the principle that within a set of professional roles, a physician assistant (PA) can be the healh care provider of choice over other health professionals. This principle may be described as "role appropriateness" and is somewhat akin to "professional specialization".

B. *Purpose of Report*

This Report documents the development of a self-assessment system by the AAPA for the continuing education of its members. The self-assessment system was envisaged as an integral coponent of a major contract supported by Health Resources Administration (DHEW) by which the competencies requisite for the entry-level generalist PA practitioner were verified, the role of the PA delineated from that of other similar health professionals, and a system developed for providing, evaluating, and accrediting continuing education programs for PAs.

The self-assessment system was, by necessity, a pilot effort since nothing like it existed for the PA profession. This is not surprising given that the first graduates from PA programs have less than five years in their practice. The self-assessment examination was eplicitly conceived in terms of a 300-item multiple-choice examination which would be carefully constructed with the help of expert committees and consultants tried out, and tentatively utilized in a model continuing education program designed to ensure professional competence among PAs.

Inasmuch as this Report highlights the processes and outcomes entailed in the development of the self-assessment system, it will opportunely be expected to serve also as its Technical Manual. The self-assessment project includes the following components: the development of test specifications based on the Role Delineation; the developoment of test items in conformance with these test specifications; the initial try-out and revision of these test items; the pilot testing of the revised test on a national sample of 100 PAs; the specification of minimum competency standards; and the development of a computer-based scoring, CME reporting and documenting system.

C. *Rationale of Project*

The self-assessment system is based upon certain axioms which are presumed self-evident. They are derived from a multidisciplinary posture formulated on the basis of experience with health professionals. These axioms will now be listed and explained so as to povide a back-drop for the project.

i) PAs are professionals and can be held responsible for their own educational maintenance and growth.

A profession is based upon service and dedication to certain human needs. Physician assistants are like other health professionals in this respect Professionals are expected to be responsible experts who are often called upon to function at the frontiers of their disciplines by using judgment and discretion in the performance of their duties. It is difficult to assume responsibility for a professional because quality service must be individualized both to the consumer's needs and to the provider's capabilities. The quality of performance may be audited by peer judgments, but such audits tend to focus on matters of gross negligence and ineptitude. The purpose of continuing education should be not merely to ensure minimally acceptable services, but rather to foster high quality health care. Continuing education should therefore be based upon felt needs, and therefore responsible self-assessment seems to provide the best answer. Mandatory programs are often doomed to become predictable failures.

ii) Given societal concerns for quality of health care and the current expectation of professional accountability, the AAPA is the most appropriate organization of the profession to assume responsibility for monitoring the quality of continuing medical education programs available and the number of credits earned by each member.

Just as individual professionals must ultimately be responsible for their own learning, so must the profession monitor itself. However, both are accountable to society and a system of verification is therefore necessary. The AAPA has developed such a CME recording system for PAs and it is planned to include self-assessment within this system.

The AAPA also enjoys distinct advantages because it speaks for the profession. The resources it enjoys go beyond membership dues, committee services, and technical input. The profession is youthful, vigorous, and enthusiastic in striving for its image and future.

> *iii) PAs are busy professionals and therefore need a continuing medical education (CME) system that is easy to access, convenient to use, self-paced, and non-threatening.*

The self-assessment idea, using evaluative examinations as the basis for CME learning prescriptions, appears sound and reasonable. This is because such examinations can be packaged so as to be convenient and inexpensive to utilize. Each PA administers such examinations to himself at his own convenience. As experience is generated by the profession, an integrated series of short examinations could be made available. In turn the PA would select units according to his professional interests and practice needs. The responses could then be scored by AAPA and learning prescriptions returned to the PA with suggestions for a variety of educational activities available to him. A PA could take parallel examinations on a certain unit several times in a year until competency is attained.

> *iv) The self-assessment programs must be practice-based and practice-oriented with emphasis on critically needed skills rather than on esoteric topics selected by teachers or indicated by recent scientific breakthroughs.*

Practitioners are interested in their day-to-day problems and look for ways to deal with them effectively. New research findings, on the other hand, clearly lack diffusion among practitioners and often remain in library shelves unapplied. One reason for this is a lack of orientation towards the practitioner in publicizing such research findings. It is difficult for a busy practitioner to derive relevance from published research. Continuing education must therefore emphasize the translation of research findings into concrete ways by which research can be applied in clinical practice.

Continuing education cannot be limited to new research findings alone. Many professionals feel the need for broadening their skills as they progress in their interdisciplinary practices. They wish to understand how other experts think and function, if only to appreciate referrals better. Some may even want to broaden the scope of their services to patients. What is needed therefore may be regular clinical skills. It is recognized that these lack the glamour of the new miracle drug or medical procedure. But to limit continuing education to the latter amounts to skirting the responsibility to enhance professional competence and thereby to ensure quality health care to society.

The question that remains is whether self-assessment examinations can be made relevant to clinical practice. The typical multiple-choice test item has tended to be frustrating because of ambiguity of the stimulus question or the trivial nature of its underlying content/ skill. This is unfortunately all too true. Good examinations are difficult to write and require an arduous process of revisions to develop. Recent successes in measurement with patient management problems (PMP) hold a distinct promise. The PMP is distinctly different from the typical multiple-choice item in that it simulates a clinical scenario and requires the making of decisions very similar to those actually made in practice.

The self-assessment examination of the AAPA uses clinicl scearios as the basis for test-items. Moreover, a unique philosophical stance was taken by requiring that all items be generated in terms of their relevance to critical skills linked to the PA role delineation. A more detailed explanation of this unique procedure will be presented later in this Report.

II. PROJECT PROCEDURES

A. *Strategy*

i) Utilize resources within profession as far as possible.

The self-assessment examination was programmed to be developed with the help of "working" committees rather than "policy-generating" committees. A "checks and balance" system was obtained by identifying two Committees, first a Test Specifications Committee of 6 persons to discuss test specifications and to develop sample clinical scenarios and test items linked to the specifications matrix; and second, an Item Writers Committee of 24 persons who worked on the actual development and revision of the test items using the test specifications. Both Committees consisted entirely of PAs and PA educators. Professional measurement consultants facilitated the process of test development by making necessary test item development, test-scoring and item analysis resources available to the AAPA. Represented on the Test Specifications Committee were the National Board of Medical Examiners and the National Commission on the Certification of Physician Assistants, both of whom have worked closely with the AAPA and the Association of Physician Assistant Programs (APAP) in the development of this program. Available for try-out of the test were members of AAPA and APAP attending the Seventh Annual Conference on Physician Assistants in Hollywood, Florida.

The utilization of PA resources in test development, aside from assuring dedication to the program by the profession, makes for needed leadership development in a young profession. Such experience is available for future capitalization and constitutes a valuable investment in the profession.

ii) Ensure long-term acceptance of program by emphasizing service to PAs.

A self-assessment program is essentially a regular service that a profession develops for its members. A successful program is based upon membership confidence in the quality of the exams; its non-punitive nature; its relevance to their professional needs; its availbility; turnaround time and cost; and, above all, the quality of the feedback provided. A service-oriented program will attend to these qualities because the target is more than the fulfillment of a governmental contract. Nothing is more aggravating to members than a central organization that seems to feed itself on short-term contracts at the cost of its membership. Loyalty and solidarity of its membership is important to the AAPA and for this reason, this self-assessment project was structured so that service would be kept in mind at all times. Everytime a PA was to be asked to provide data by responding to a try-out version of the examination, the Committees asked themselves: What benefits can we provide the PA in return?

B. Work Plan

1. Contractual framework: The revised contract (February 9, 1979) provided that a criterion-referenced self-assessment examination be developed by utilizing the following critical steps:

i) Obtain services of consultants with expertise in development of criterion-referenced self-assessment tools.

ii) Select competency areas from the major responsibility domains of the entry-level generalist physician assistant, using additional criteria in the process of selecting topics for the self-assessment tool.

iii) Using a Working Committee, establish the test descriptive scheme and generate items for the self-assessmen tool.

iv) State the test's descriptive scheme which constitutes the self-assessment tool and identify the specific items.

v) Develop (includes testing) the self-assessment tool. Pilot test the exam on 100 PAs.

vi) Submit draft self-assessment tool and a description of pilot testing results for review and approval by Project Officer.

vii) Design a program by which the self-assessment tool is made available to physician assistants.

viii) Prepare draft final report on the self-assessment tool including recommendations for its future use on profiling practitioners in the field and development of learning packages.

Earlier, the AAPA had proposed a twelve-step scheme by which it indicated that the test specifications would be developed by a Test Specifications Committee of six experts in criterion-referenced testing and including representatives from the National Commission on Certification of Physician Assistants (NCCPA) and the National Board Medical Examiners (NBME). The item development would be undertaken by content specialists at two workshops, the first of which would instruct them in the development of such test items. It was also hoped to be able to get content experts in the field to write items with the help of written instructions and the test specifications. The Test Specifications Committee would then meet to review the test items and assemble the examination. The exam would then be pilot-tested, results provided to the PAs, and the Test Specifications Committee convened a third time to make recommendations for future use of the self-assessment examination.

2. The revised Work Plan: Early in March 1979, the AAPA hired Ayres D'Costa, Ph.D., Associate Professor of Health Professions Education at The Ohio State University, to serve as the Consultant for the Project.

After some initial discussions among the project staff, the Consultant, and the Project Officer, a Schedule for Test Development was agreed upon. (See Table 1). This Schedule recognized the need for an additional try-out of the test items being developed. This try-out was scheduled for April 26, 1979 during the AAPA-APAP Convention in Hollywood, Florida. Working around this fixed date, the first Item Writers' Workshop was utilized to develop items and the second Item Writers' Workshop was scheduled to revise the items on the basis of the item analysis data and the comments received from PAs. All other aspects of the contractual framework were left intact.

C. *An overview of the Actual Work Schedule:*

Given a Test Specifications Committee (TSC) and an Item Writers Committee (IWC), each of whom would meet twice during the project period, it was decided to bring the Test Specifications Committee together early in April and once again towards the end of the project. All Committee meetings were called Workshops and became intensive work-sessions designed to produce specified project products.

The first TSC Workshop resulted in 1) prioritized lists of research and program objectives for the Self-Assessment Examination, 2) a test specifications matrix using the 11 role delineation areas along one axis and 5 skills levels along the other axis, 3) samples of test items generated from Critical Incidents/Scenarios, and 4) a Revised Schedule of activities planned for the project. The next landmark event was the Item Writers Workshop on April 16-18,

TABLE 1

REVISED SCHEDULE FOR TEST DEVELOPMENT

March 24	Orientation materials sent to Test Specifications Committee
April 3-5	Test Specifications Committee Meeting #1 meets with project staff and consultant. Test objectives and specifications developed.
April 8	Orientation materials sent to item writers, both workshop participants and PAs who will develop test items in their practice settings. Both groups will be asked to write 10-15 test items. The practice group will mail these to the National Office prior to May 1. The workshop group will bring these questions with them.
April 16-18	Item Writers Workshop #1 with project staff and consultants. Items will be written and reviewed. Each participant is expected to develop about 10-15 items during the Workshop.
April 26	About 8 to 10 (non-parallel) test forms with about 30-40 items each will be tried out at the PA Convention in Hollywood, Florida. Matrix Sampling Approach will be used. The PAs will review each item for readability, social desirability, and relevance to practice.
April 27-May 2	Consultants will review item analyses data, as well as PA comments, to perform some preliminary revisions of items. The revisions and item analyses data will be sent to each item writer in order to request additional revisions based on medical content. Some items will need to be dropped, new ones developed, most will be revised.
May 17-19	Item Writers Workshop #2. Discussion of proposed revisions for items; needs with respect to items; overall test quality. Develop instructions for test administration, scoring and interpretation strategy.
May 20-21	Consultants prepare final test form for mailing as trial self-assessment instrument to 100 PAs,
May 25	Self-assessment instrument (Trial Form) mailed to 100 PAs.
June 4-13	PAs return completed self-assessment examination to Consultant in self-addressed envelope.
June 14-16	Consultants score and item analyze self-assessment exam.
June 17-18	Consultants review item analysis, make necessary revisions and prepare report.
June 21-23	Test Specifications Committee Meeting #2. Exam and report reviewed. Comments and further action suggestions recorded.
June 29	Project Final Report due to HEW Project Officer.

1979. Prior to coming together, the members of this Committee received instructions for writing scenarios and test items, sample test items and scenarios, a project Schedule, and the PA Role Delineation. They were asked to identify critical incidents related to the Role Delineation and to bring these along to the "Workshop". The Workshop began with an overview of how test items are written to test specifications and revised on the basis of item analysis. A *Guide for Item Writers* was prepared with sections on Item Styles, Item Editing Principles, Item Revision Principles Based on Item Analysis Data, Writing Test Items on Interpersonal Skills, and Some Basic Concepts on Bloom's Taxonomy. The Test Specifications were discussed and the need to develop scales pointed out.

The 24 persons attending this IWC Workshop worked in four groups and produced four 80-item Test Sections. Each test item was referenced to the Role Delineation, to a Scenario, and ultimately to the Test Specifications Matrix.

These four Test Sections were edited and "tried out" on PAs attending the AAPA-APAP Conference. This first Try-Out consisted of responding to the 80 items, then rating each item for relevance to PA practice, and finally indicating any problem words/phrases in the test-items. The respondents remained anonymous and no feedback was promised other than the Answer Key.

The responses received for this try-out were computer scored, item analyzed, and frequency distributions and other statistics generated. The relevance ratings were likewise scored and item analyzed. All his data was then summarized and mailed to the respective group of Item Writers responsible for the Test Section. Written and other comments on the test and individual items were also summarized.

Members of the Item Writers Committee were urged to use the above material in preparing for the second Workshop on May 17-19. Specifically, five tasks were identified as homework: completing the Scenarios file, verifying the allocation of the items to the test specifications matrix, revising items using the item analysis and relevance summary data, developing new test items where needed to fulfill the specifications matrix, and reviewing of test item options to generate appropriate feedback on error patterns of responders.

At this point, it is necessary to mention that test items were also written by some field writers (PAs and PA educators selected by AAPA staff) using the *Guide for Item Writers* and other written materials available at this point in the project. Unlike the Item Writers Committee, the field writers worked on their own at home. A fifth Test Section of 80 items was thus developed. Section 5 was administered to a group of PAs and the responses were scored and item analyzed.

Several PA programs responded to the AAPA call for Test Items. A large number of test items was thus accumulated. These items are of variable quality and have not been critiqued nor coded to the Specifications Matrix.

The pace of the second Workshop for Item Writers was hectic but a considerable amount of time was spent in reviewing the feedback capabilities of the Examination. The Item Writers recommended that it would be more meaningful to PAs if additional scales were developed using *Body Systems and Medical Intervention Type* as the basis. This resulted in a set of 28 scales. All available test items were classified in terms of these 2 new criteria--Body Systems, Medical Intervention Type, as well as the original test specifications criteria--Role

Areas and Skill Levels. Additionally, three other criteria were utilized for analyzing the test items, namely, patient age, medical specialty, and common disease categories as identified by the Medical College of Virginia (Marsland et al., 1976). This effort resulted in a test item bank of about 425 items with all items classified by these seven categories. The correct answer for each item was also documented in terms of standard medical texts.

The second Item Writers Workshop resulted in 315 items. These were assembled into two Sections, with 160 and 155 items respectively, in order to fit a standard Digitek answer sheet. A third Section was added to obtain data on the practice profile of the PA taking the self-assessment examination, and also to ascertain felt continuing education needs in terms of the 28 System Scales. A few additional questions were added to get the professional background of the PA and to receive evaluative ratings on the project from the PA.

Early in June, a self-assessment package, consisting of the three Sections with appropriate answer sheets and directions for self-administration and use of return envelope, was mailed to a random sample of 274 PAs. This constituted the second try-out of the Exam.

As scheduled, the second Test Specifications Committee Workshop was held on June 21-23. At this point, usable responses had been received from more than 100 PAs, the number that had been originally planned for. Several tasks remained before these responses could be scored and reported on. These were: 1) the verification of the correct response to each test item by this independent other Committee, 2) the verification of the 28 System Scales and of the classification of the test items in terms of these scales, 3) the development of appropriate Role Scales and the verification of the classification of the test items in terms of these scales, 4) an independent review of each test item in terms of its quality, 5) the determination of the scores expected (Nedelsky Method) of a minimally competent PA, 6) a review of the four-page Report to be computer-generated and provided to each PA taking the self-assessment examination, and 7) a list of recommendations for additional work and next steps with this project.

The Test Specifications Matrix was redefined in terms of 17 Role Scales based upon a regrouping of the 11 areas in the Role Delineation, a revision of the five skill levels, and the introduction of Body Systems categories into the Role Delineation. Revisions were recommended to about 60 of the items, and several were tagged for deletion from the examination. Unfortunately the Committee did not have access to the Item Analysis on the Revised Examination at the time of its meeting (the responses had barely been received then) and so the recommendations were entirely judgmental.

Individual comments and extensive reviews of test items have since also been received from PAs in the field. The examination will therefore need to be thoroughly revised on the basis of all these comments and reviews, as well as on the basis of the item analysis data now available. An initial review of these data by the Consultant was used to arrive at tentative decisions on the scoring key for this try-out reporting.

The four-page Report had been computerized and an *Interpretive Leaflet* prepared to accompany this Report to each of the 112 PAs who participated in this Second Try-Out of the examination. A special computer system has been developed to generate these Reports and to provide the AAPA with: 1) the usual measurement quality indices for this examination such as reliability, coefficient of agreement and standard error of measurement for each scale, 2) a summary of the scale scores for the total group in terms of mean, standard deviation, range and frequency distribution, 3) a summary of the practice profile scores and continuing education needs scores for the total group, and 4) a summary of the evaluative feedback provided by the 112 PAs on the self-assessment project.

D. *List of Products Developed/Under Development*

Several products have been generated by the project. Those enclosed with this Final Report are indicated by asterisk. Intermediate products and by-products are listed but not enclosed. Products that are under development are indicated by italics.

* i) Unprioritized List of Program and Research Objectives for the Self-Assessment Examination (Exhibit A)

* ii) The Test Specifications Matrix and its proposed Implementation Chart (Figure 2)

* iii) Instructions for writing a Scenario and a sample test item generated from a Scenario (Tables 7, 8, and 9)

iv) Guide for Item-Writers

v) Five 80-item Test Sections (First Try-Out)

vi) Item Analysis Results for the five Test Sections with usual scores statistics

vii) Frequency Distributions of the Relevance Ratings for the four Test Sections

viii) Summary of Written Comments on items in the four Test Sections

ix) A Scenarios File listing critical incidents for the Role Delineation

x) Test Items (Unclassified)

xi) Test Items File for Classified Items

* xii) The 28 Body System X Medical Intervention Scales (Figure 4)

* xiii) The 17 Role Area X Body System Scales (Figure 3)

xiv) Correct Answer Documentation File (to be merged into Test Items Bank)

xv) The three Sections of the Self-Assessment Package (Second Try-Out)

* xvi) Scores expected of Minimally Competent PA (Nedelsky Method) by Scale (Exhibits G, H)

* xvii) The Individualized Report (Exhibit J)

*xviii) The Interpretive Leaflet (accompanies Individualized Report) (Exhibit K)

* xix) Statistical Summary Reports on Scores (Exhibits C, D, E)

* xx) Summary of Evaluative Ratings of Project (Table 10)

xxi) Computer Scoring and Reporting System

* xxii) List of Recommendations for Future Efforts (Chapter IV)

xxiii) Item Analysis Results for the Second Try-Out with usual scores statistics

xxiv) Research studies/papers

xxv) Symposium for presentation at the 1980 Annual Convention of the American Educational Research Association

xxvi) The training of a small group of PAs in the technical aspects of item writing, item revision, and test development procedures

xxvii) Listing of computer cards documenting characteristics of all items in Test Item Bank

E. *Project Problems Encountered*

1) The time crunch. For several reasons, the project did not get actively underway until early March 1979. Therefore, the process for developing the products to this project had to be compressed. The AAPA was fortunate to receive a 45-day extension from HRA so that the essential products could be completed as proposed.

2) Quality of Test Items. It was difficult to produce test items in larger quantity and better quality despite the excellent efforts on the part of all concerned, because time is needed to train more physician assistants in the technical aspects of item writing and item revision. A few physician assistants are currently available with such expertise but their number is not large enough because the profession is young. The item-styles utilized in the examination, the quality of the response options, and the cognitive level of the questions can be improved as more time becomes available and experience is gained.

3) Technical Problems. Although these will be discussed in greater detail in another section of this Report, the project had to contend with the current deficiencies in the technical state-of-the-art relative to self-assessment methodology, the measurement of professional competency, the setting of minimum competency standards, and the development of criterion-referenced examinations. Traditional testing, as contrasted with self-assessment, uses rigorous test administration procedures. Little seems to be known about self-assessment, and even less about why and how professionals seek continuing education. Professional competence remains a complex set of skills, the most critical of which, such as interpersonal and attitudinal skills, are still very difficult to measure by multiple-choice examinations. The techniques for setting minimum standards are typically judgmental and are therefore prone to error and bias. The situation with criterion-referenced testing is like that of the tail wagging the dog. The public is sold on the idea, but the technical cupboard is yet bare. The techniques available for the development of such examinations are yet on the frontiers of measurement technology and therefore not easily available.

III. TECHNICAL ISSUES

A. *Objectives of Self-Assessment Examination*

A self-assessment, unlike a self-rating, does not necessitate a self-indictment. Ratings seem to have an end-point finality about them that influences the manner in which individuals are willing to look at themselves. Perry (1977) noted that although physicians are very happy with physician assistants, the validity of self-ratings of performance by physician assistants was generally questionable. Futhermore, Kegel-Floom's research quoted by Perry indicated that personality characteristics substantially bias self-evaluation of performance.

A self-assessment is an opportunity for self-improvement without any judgmental labels or punitive consequences. It is likely that the professional's interest in seeking continuing education is influenced by his feelings of inadequacy or his need for better knowledge and skills. Self-assessment could be an aid to kindle such feelings or needs. The fact that travelers will test their IQ during their leisure times may indicate an innate human curiosity about oneself and augurs well for the practical utility of making self-assessment tests available on a voluntary basis.

Research on self-evaluations of dentists (Milgrom et al., 1978) indicated that their accuracy increased as they became more specific. In other words, professionals are more threatened by global assessments and more willing to acknowledge deficiencies in certain specific aspects.

It is also essential to emphasize the diagnostic intent of self-assessments in order to differentiate them from certifying exams (Engel, 1976). These differences significantly impact upon the manner of constructing and interpreting such tests.

The ultimate purpose of the self-assessment examination is to enable physician assistants to maintain their competence and thereby ensure the quality of health care. A competent professional may be defined as one who knows how to do well the job expected of him and is able to translate this knowledge into his practice. Competence includes knowledge, application, and attitudinal skills. At a higher-level, application develops into technical problem-solving as well as into interpersonal communication skills. As a result, one might define, using a combination of Bloom's cognitive system, Krathwohl's affective system, and Gagne's learning model, a five-level system of competence defined as follows: knowledge, application, technical problem-solving, interpersonal communications, and professional attitudes.

The key to the maintenance of professional competence does not lie merely in the providing of appropriate continuing education programs. Professionals tend to be busy persons who are not easily convinced of the practical utility of attending educational programs. Often such programs are not targetted to their immediate needs, or they are inconvenient to attend, or they are inappropriately handled by instructors, or they remain unknown to the busy professional.

Recent efforts by the professions to require recertification on a regular basis by their membership are based upon the rise of malpractice suits and a continuing demand by an increasingly better-educated society for good quality care. Professional accountability, however, is limited to what one claims to be able to do and is actually engaged in doing. The maintenance of competence is therefore circumscribed by what a professional professes to be able to do by virtue of his role and by what he actually deals with in practice. One way of looking at professional accountability is via a consumer-provider model of professional roles (D'Costa, 1975) shown in Figure 1. There are four forces at play: the role expectations of the profession, the role expectations of the patient and consumers, and those of the individual professional himself.

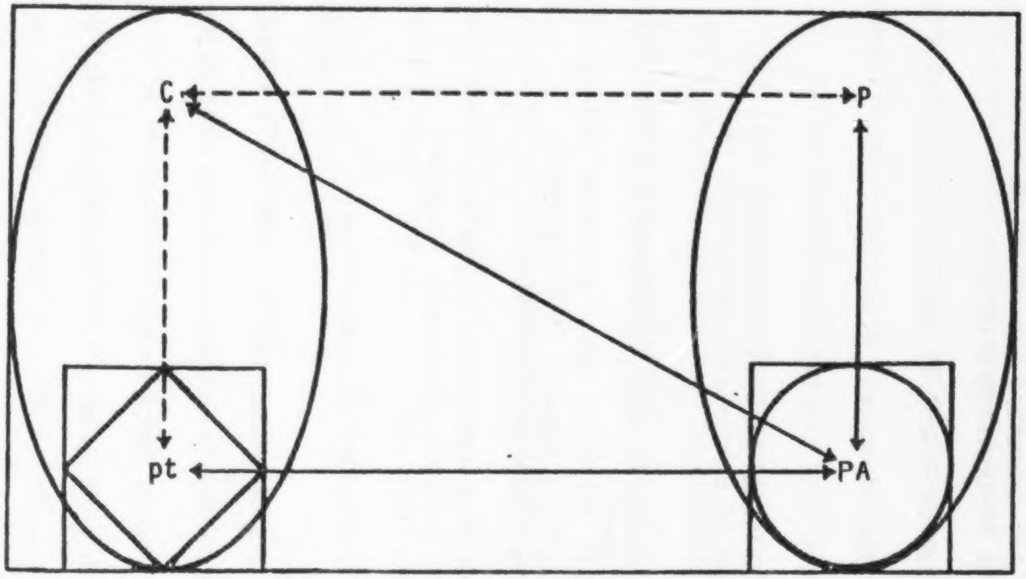
The Self-Assessment Examination of the AAPA was accordingly designed to encourage a PA to plan his continuing medical education (CME) in terms of these various forces, namely: professional role, practice expectations, and individual felt needs. The AAPA sees as its role the development of appropriate self-assessment tools, the facilitation of such planning, and the providing of worthwhile continuing education programs on an efficient basis.

In assessing this AAPA role at the first Technical Specifications Committee Workshop, distinction was made between day-to-day programmatic goals and technical/research goals of a self-assessment examination (Exhibit A). The five most important program-related goals were identified as:

1) Develop a national profile of PA-CME needs;
2) Ensure that the self-assessment was not narrowly conceived as an aid to recertification but rather as an aid to the maintenance of professional competence and quality of practice;
3) Recognize that, since the primary purpose is to help the PA plan his continuing medical education, the self-assessment program should stress sufficient feedback to the PA;

FIGURE 1

THE FOUR TYPES OF EXPECTATIONS OF HEALTH PROFESSIONALS



Codes

P = Providers of health care/profession

-> = Process

C = Consumer of health care/society

pt = Patient

|_| = Structure

4) Recognize that the present paper and pencil examination may not encompass all aspects of clinical competence; and

5) Recognize that this self-assessment is geared to individual needs and therefore may not be directly useful to evaluate PA training programs.

At the technical/research level, it was recognized that the state-of-the-art is far from adequate. Accordingly the following were identified as the five most important technical goals:

1) Define the core/critical skills, behaviors, and knowledge expected of entry-level generalist PA professionals;

2) Study the relationship between competence and tasks frequently done;

3) Study the relationships between self-expressed competence and test-derived competence;

4) Identify strengths and weaknesses of PAs in terms of training program, geographic location, and practice specialty; and

5) Identify causal dimensions of professional performance.

On second thought, the Committee decided that while these goals were good to maintain for perspective purposes, the major efforts of this project should focus on the development of the examination and on the setting up of a self-assessment model with emphasis on feedback.

B. Technical Rationale of Test

The original HRA contract called for several self-assessment tools each with a correlated individual independent study package. As discussions between the AAPA and HRA continued it became evident that it would be too early to embark upon such a massive program. Accordingly, the contract was modified to specify the development of a single criterion-referenced self-assessment tool. Not only was this goal reasonable in the circumstances, but it also provided opportunity for the development of the necessary technical framework upon which a system of self-assessment tools could be developed in the future. Merely generating several self-assessment tools might have been disastrous.

The main reason for the above line of thinking lies in a basic principle of criterion-referenced testing. Popham (1978) states that such tests are designed to ascertain an individual's status with respect to a well-defined behavior domain. The precise definition of the domain in terms of skills is critical to the concept of criterion-referenced testing because of the need to make generalizations about the mastery or non-mastery of these skills based on test scores.

C. *Test Specifications*

Hambleton and Eignor (1979) provide a 12-step process for developing and validating criterion-referenced tests. Unfortunately, their emphasis is on objectives and the specification of item formats and number rather than on the crucial matter of domain definition advocated by Millman (1974). Merely listing objectives related to criteria becomes an atomistic approach that is limited in meaningfulness and relevance when it comes to interpretation or self-assessment. To this project the quality of self-assessment is paramount and for this reason the matter of domain definition becomes very important (Pottinger, 1977). The criteria or objectives must be linked to the main domain and the linkages must be clear.

Only then will a PA recognize the implications of his weakness in some skills in relation to his overall performance as a PA.

The domain of this self-assessment examination is the performance expected of a minimally competent generalist PA. Wilson (1976) argues that competency assurance in a credentialing program "must be based on a sound generic position classification". Fortunately, such an analysis has been completed and verified in the case of the PA profession by the AAPA. Indeed this project is an integral component of that major effort.

The role delineation of a profession describes the tasks which a practitioner must be competent to perform. This contrasts with other approaches, such as task inventories, which list all tasks that a practitioner can, should, or might perform. The role delineation thus provides a position classification and is a minimum standard expected of all practitioners in the profession. A role delineation is expressed in terms of performance responsibilities rather than just knowledge expected.

The 1979 version of the *Role Delineation* for the PA (see Volume II) lists 11 major areas of responsibility (Exhibit B). Each area is extensively defined in terms of specific responsibilities. Together, the major and specific responsibilities define the domain of this self-assessment exam.

The structure of the domain was initially recognized as the 11 areas of major responsibility. The specific responsibilities under each area were also recognized for purposes of definition and item generation, thereby ensuring fidelity to the meaning assigned to each role area in the *Role Delineation*. However, their number was considered too numerous to include in a test specifications matrix.

The 11 areas of responsibility served as the major content areas defined along one dimension of a specifications matrix. It is customary to identify skills levels as the other dimension. Typically, Bloom's taxonomy (1956) of cognitive skills has served as this dimension. In the present situation a five-level scheme generated as follows: *knowledge, application, problem-solving, interpersonal skills, and professional attitudes.*

A 11 X 5 matrix thus served as the initial test specifications matrix. It was recognized that the 55 cells in this matrix were too many to utilize as scales for feedback purposes. Concern was expressed by the Test Specifications Committee about the reliability of test items related to role areas, such as: recognize interdependent relationship, demonstrate professional behavior, promote acceptance of PA role, and maintain competency. It was also noted that measurement techniques available for attitudinal and interpersonal skills are not of the usual paper-and-pencil type.

It was the intent of this project to develop a test specifications matrix that represented the ideal expectations of a self-assessment examination and to use this as the target during the item development process. However, in the implementation of this project, this was found difficult to implement and a compromise procedure was utilized. The Committee members began by assigning ideal weights for the specifications matrix on an individual basis, but later during the group discussion process they negotiated compromise weights with each other using current measurement realities as their basis. Table 2 presents the weights (shown within boxes) arrived at by the Committee for the row and column totals or matrix

TABLE 2

TEST SPECIFICATIONS MATRIX CELLS AND MARGINALS

Skill Area	1	2	3	4	5	
1	0.78	1.8	2.7	0.6	0.12	6
2	0.39	0.9	1.35	0.3	0.06	3
3	3.9	9.0	13.5	3.0	0.6	30
4	5.85	13.5	20.25	4.5	0.9	45
5	6.24	14.4	21.6	4.8	0.96	48
6	6.24	14.4	21.6	4.8	0.96	48
7	5.85	13.5	20.25	4.5	0.9	45
8	5.85	13.5	20.25	4.5	0.9	45
9	1.95	4.5	6.75	1.5	0.3	15
10	0.78	1.8	2.7	0.6	0.12	6
11	1.17	2.7	4.05	0.9	0.18	9
	39	90	135	30	6	300

marginals. Low weights were assigned to Areas 1, 2, 10, and 11 and to Skill 5 even though these weights did not reflect their importance to competent performance. The weights assume that the total number of test items would be 300.

The derivation of individual cell weights was initially done mathematically, using an expected frequency computational approach. Table 2 reflects such expected values. However, an actual specifications matrix does not need to have each cell weight proportionate to its respective marginals (row and column totals). Instead some cells can be left blank and others enhanced (in order to reflect the real world) without violating the marginals. Such a refinement of the Test Specifications Matrix is presented in Figure 2.

The advantage with this scheme in implementing test specifications lies in its feedback capabilities. Instead of 55 cells some of which would have very few test items allocated, it now becomes possible to cluster test items around relevant cells/scales. Note, too, that the number of scales can be reduced to a manageable number.

The boxes in Figure 2 represent six items each. This was done because research by Eignor and Hambleton (1979) on effects of test length on selected test score reliability and validity indices indicated that, depending upon the domain characteristics and the decision-making strategy used, even tests with as few as 6 items could be effective in criterion-referenced testing. It is assumed that two or more boxes can be combined to form a single scale wherever appropriate. However, the potential for creating more than one scale within each cell permits other criteria to be recognized thereby acknowledging the multidimensionality of the test domain. Note that this imple-

FIGURE 2

IMPLEMENTING THE TEST SPECIFICATIONS MATRIX USING TS COMMITTEE MARGINALS

Skill / Role Model Area	1. Knowledge	2. Application	3. Problem Solving	4. Inter-Personal Skills	5. Professional Attitudes	TS Committee Totals
I. Recognize Interdependent Variables			⧄		⧄	6
II. Demonstrate PA Professional Behavior		⧄			⧄	3
III. Promote Preventive Health Care		☐☐	☐☐	⧄☐	⧄	30
IV. Establish Health Status Data Base	☐	☐☐	☐☐☐☐	⧄		45
V. Analyze Data Base	☐	☐☐☐	☐☐☐☐		⧄	48
VI. Formulate Health Mgt. Plan	☐	☐☐☐	☐☐☐☐			48
VII. Implement Health Mgt. Plan	☐	☐☐	☐☐☐	☐	⧄	45
VIII. Monitor Health Mgt. Plan	☐	☐☐	☐☐☐	⧄		45
IX. Establish Effective I-P Relationships	☐	⧄	⧄	⧄☐	⧄	15
X. Maintain PA Competency	⧄		⧄	☐☐		6
XI. Promote Acceptance of PA Role			⧄	☐	⧄	9
TS Committee Totals	39	90	135	30	6	300

Symbols: ☐ Scale of 6 items ⧄ Recommended/Part Scale mentation of the Specification Matrix does not change the originally prescribed matrix marginals. The items represented by the total number of boxes add up to the row and column marginals/totals.

The allocation of the boxes (potential scales) to the cells in the matrix was done so as to make optimum practical sense given the nature of the role responsibilities, the level of skills required, and the recommendations of the Test Specifications Committee.

D. *Scales Definition*

The theoretical or *a priori* derivation of scales assumes that enough is known about the real world of PA competence by the Specifications Committee. This was not the case in this project primarily because the profession is young and little data is currently available. For these reasons a "successive approximations" strategy was employed. The issue of scales definition was taken up at every Workshop of the two Committees and it was not until the final Workshop that the scales were finalized for this project's purposes. In keeping with this strategy it can be expected that the future will see additional modifications to the two sets of scales currently defined by the project for the examination.

The two sets of scales are named: "Role" and "Body System" (see Figures 3 and 4). In actuality each set of scales is defined by a matrix with two criteria. The 17 Role scales are defined by the cells of a matrix obtained from the 11 role areas and 13 body systems. The 28 Body System scales are defined from the matrix defined by 13 body systems and four medical intervention types. Table 3 provides broad descriptions of the Role Area Scales. Several questions arise at this point. How was "Body Systems"

TABLE 3

DESCRIPTION OF THE 17 PA ROLE AREA SCALES

Scale	
1. Professional Role	Recognize interdependent relationship with supervising physician Maintain competency Promote acceptance of the role Items on this scale are related to understanding the PA role, working within the role, maintaining competency as a PA, explaining the PA role to others, and displaying appropriate PA behaviors.
2. Interpersonal Behavior	Demonstrate professional behavior Establish effective interpersonal relationships with patients, professionals, and others. Items describe behavior which involves interactions with others, especially to demonstrate concern, respect, and empathy with the other.
3. Gather Data - Resp./CV 4. - GI/GU 5. - Psy/Neuro	Establish health status data base Items demonstrate basic knowledge essential to the data gathering process, i.e., the PA knows what information to collect and which diagnoses are possible, given certain information.
6. Analyze Data - Endo. 7. - Resp/CV 8. - Hema. 9. - GI/GU 10. - Psy/Neuro	Analyze health status data base These items demonstrate the use of knowledge in the decision-making process, i.e., the PA can interpret data from laboratory tests, history, and physician examination to lead to a working diagnosis.
13. Manage Patients - Resp/CV 14. - GI/GU 15. - Repro. 16. - Psy/Neuro	Formulate health management plan Implement health management plan Monitor health management plan Items demonstrate whether, for a diagnosed problem, the PA can develop a plan of action, carry out the plan, and/or monitor progress in order to make any necessary modifications in the plan.

Note: Scale 11 is Formulate Plan - Musc/Skel
Scale 12 is Handle Derm Problems
Scale 17 is Handle Pharm Problems

These three scales combine two or more of above descriptions.

FIGURE 3

MAP OF THE 17 ROLE AREAS X BODY SYSTEM SCALES

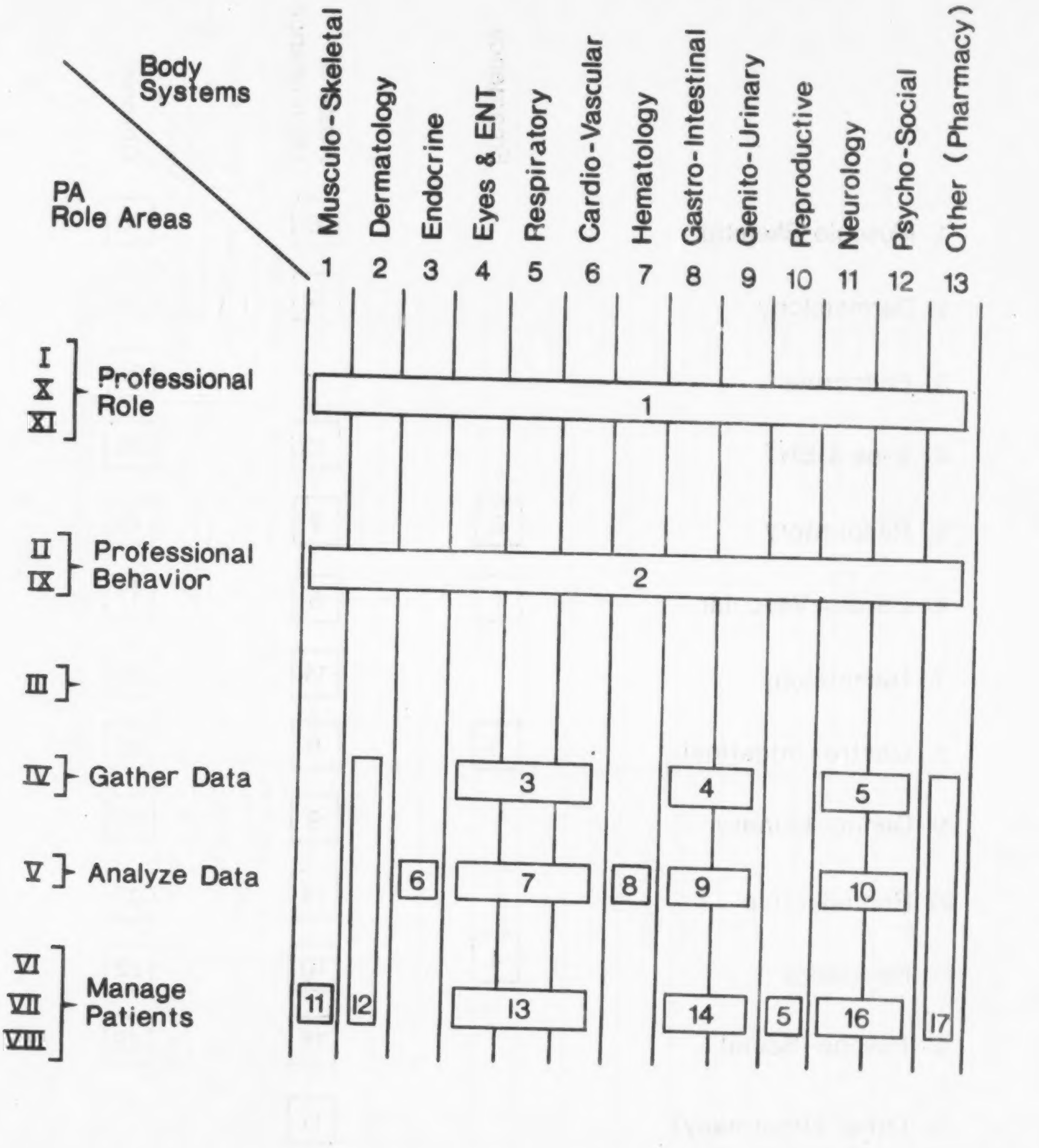

Body Systems
PA Role Areas
1 Musculo-Skeletal
2 Dermatology
3 Endocrine
4 Eyes & ENT
5 Respiratory
6 Cardio-Vascular
7 Hematology
8 Gastro-Intestinal
9 Genito-Urinary
10 Reproductive
11 Neurology
12 Psycho-Social
13 Other (Pharmacy)
I X XI Professional Role
II IX Professional Behavior
III
IV Gather Data
V Analyze Data
VI VII VIII Manage Patients
1
2
3
4
5
6
7
8
9
10
11
12
13
14
5
16
17

FIGURE 4

MAP OF THE 28 BODY SYSTEM x MEDICAL INTERVENTION SCALES

	Emergency	Acute & Maintenance	Chronic
1. Musculo-Skeletal		6	18
2. Dermatology		12	24
3. Endocrine			23
4. Eyes & ENT		13	25
5. Respiratory	2	7	19
6. Cardio-Vascular	1	5	17
7. Hematology		14	26
8. Gastro-Intestinal	3	8	20
9. Genito-Urinary		9	21
10. Reproductive		15	27
11. Neurology	4	10	22
12. Psycho-Social		16	28
13. Other (Pharmacy)		11	

TABLE 4

CODES/LEVELS FOR CLASSIFICATION VARIABLES

Code #	I Skill Level	II Role Del. Model	III Patient Age	IV Intervention Type	V MCV Disease	VI Body System	VII Medical Specialty
1	Knowledge	MD Interdependence	Pediatric	Acute	Maintenance	Musculo-Skel	Int. Medicine
2	Problem Solving	Profl. Behavior	Young Adult	Maintenance	Upper Resp. Inf.	Dermatology	Surgery
3	Interpersonal	Prevention	Adult	Emergency	Hypertension	Endocrine	Inf. Diseases
4		Establish Data	Geriatric	Chronic	Depression	ENT	Ob-Gyn
5		Analyze Data			Arteriosclerosis	Respiratory	Pediatrics
6		Formulate Plan			Diabetes Mell	Cardio Vascular	Psychiatry
7		Implement Plan			Arthritis	Hermatology	Radiology
8		Monitor Plan			Genito Urinary	GI	Opthalmology
9		Interpersonal			Obesity	GU	Pathology
10		Competency			Otitis Media	Reproductive	Pharmacology
11		Acceptance of Role			Peptic Ulcers	Neurology	Physiology
12					Vulvo Vaginitis	PsychoSocial	Preventive
13					Headache	Other	
14					Anemias		
15					Cong. Heart Fail		
16					Other		

selected as a criterion? Why "medical intervention type" and why not some other criterion such as "patient type"? Why is "Body System" utilized a second time with the 11 "Role areas?" What happened to the five "Skills levels?" What types of criteria were considered before these decisions were made? These considerations are critical to an understanding of the test domain defined and to an appreciation of the problems inherent in developing useful test specifications for a self assessment examination.

To begin with, it was understood that the starting point for the test would be the *Role Delineation* for PAs. Considerable effort had gone into the development of the *Role Delineation* and into its verification. It was also recognized that the 11 Role Areas, although judged critical to PA competence, may not serve as the best feedback mechanisms for continuing education purposes. For example, being told that one is deficient in data gathering skills may be too global a diagnosis in terms of making a meaningful remedial prescription understandable to a PA. The Item Writers Committee was particularly sensitive to this dilemma and urged consideration of other criteria, especially Body Systems, on an additional basis. The Test Specifications Committee was sensitive to this problem too, and had recommended that other criteria, such as patient type, medical intervention type, body systems, medical specialty, common patient presenting symptoms (MCV Disease) be also considered when developing test items. The intent on their part was representation of clinical practice. Table 4 presents the levels for each of the seven classification variables.

It was the judgment of the two Committees that Body Systems represented the most useful criterion to use in scale development for several reasons: 1) most text books are organized by body systems, 2) body systems provide a better reference approach in studying patient problems, 3) medical specialty is not useful to physician assistants because of the profession's emphasis on primary care, 4) the most commonly presented patient symptoms (MCV Disease Categories, Marsland et al., 1976) are not comprehensive and are inconvenient because there are too many categories.

"Medical intervention" was selected over "patient characteristics" because it provides a well-defined classification scheme in patient care. Emergency care is now well recognized as a class by itself and health maintenance is fast emerging as a new thrust of societal interest.

The five skills areas were very much in the minds of item writers when developing the test. However, the number of levels was dropped from five to three in order to simplify the task for item writers. Note the special instructions (Table 5) to help item writers recognize these three levels of skills. Furthermore, test items in each Test Section were classified in terms of the original Test Specifications Matrix as shown in Table 6. This three level approach to skills is the one that has been adopted for all test items in the Item Bank now developing with AAPA. The concern for representing all three types of skills in each scale of the test persisted throughout the project. It was recognized that interpersonal skills are the most difficult to measure, and that most test items tend to become of the knowledge level.

E. *Item Generation*

The major test development approach utilized in this project was derived from the critical incidents technique first proposed by Flanagan (1954). Given the 55 cells defined by the 11 role areas and the five skills levels, item writers were asked to identify critical incidents for each cell. Furthermore, the item writers were asked to utilize their experience to describe the critical role of the PA in the incident (patient scenario) in terms of skills needed and errors likely. As critical incidents were identified and the needed major skills and typical errors noted, items began to be written and situational details added on. Discussions ensued within each group as to how typical a given scenario was in PA practice and changes were accordingly made. This approach to item generation used in conjunction with the Role Delineation is unique in that it maximizes a concern for the critical characteristics of job performers rather than merely considering critical dimensions of the job (Pottinger, 1977). Table 7 presents the instructions to Item Writers. Table 8 presents the Critical Incident/Scenario developed by one item writer for Role Area 4. Table 9 presents a Test Item generated from this Scenario. Note that the Guide for Generating Scenarios bypasses the usual development of behavioral objectives and amplified objectives as recommended by Popham (1978). Instead the item writer moves directly to the identification of a critical incident related to the test domain when the PA either functioned very effectively or very ineffectively. The second step involves the usual amplification process (conditions, skills, errors), but it is modified so as to obtain material needed to

TABLE 5

HOW TO CLASSIFY TEST ITEMS BY SKILLS LEVEL[1]

In classifying each item, three skills levels were used:

- knowledge
- problem solving
- interpersonal skills

1. *Knowledge* refers to any item requiring factual recall of information This was used in cases where a diagnosis (as uncomplicated trichomonal vaginitis) or condition (as dark urine) was identified and specific treatment procedures, data gathering techniques, or potential causes were requested. The key element in these items is that the examinee is given a clearly identified and limited context in which to provide specific information (lab procedures for vaginal discharge are...; conditions causing asthmatic symptoms in the pediatric age group include...)

2. *Problem solving* refers to any item involving two steps. First, the examinee must analyze and order the information provided in a problem situation (logical thinking). In this first step, the examinee infers what the problem really is. Second, the examinee both recalls and applies previous knowledge and experience in determining appropriate courses of action. The category problem solving was used primarily in those items describing a patient with signs, symptoms and/or presenting complaint. These items usually required both identification of the problem and determination of appropriate actions.

3. *Interpersonal skills* refers to those items clearly requiring the use of effective human relations skills.

---

[1] This classification strategy was prepared by Dr. Cherry Turner, Division of Medical Education Research and Evaluation, College of Medicine, The Ohio State University.

TABLE 6

ASSIGNMENT OF TEST SECTION 1 ITEMS TO SPECIFICATIONS MATRIX

ROLE MODEL AREA	KNOWLEDGE	PROBLEM SOLVING	INTERPERSONAL SKILLS
I. INTERDEPENDENT RELATIONSHIP		30	32
II. PROFESSIONAL-BEHAVIOR			72
III. PREVENTIVE HEALTH CARE	62	42,43	
IV. ESTABLISH DATA BASE	25, 53, 55, 56, 60 76, 79	4, 7, 9, 12, 13, 19, 27, 28, 39, 52, 54, 65	
V. ANALYZE DATA BASE	15, 26, 34, 35, 47 61, 64, 70, 78	3, 22, 24, 29, 31, 40, 45, 46, 48, 49, 50, 59, 63, 68, 69, 71, 80	17
VI. FORMULATE PLAN	2, 57, 58	37, 41, 74, 77	
VII. IMPLEMENT PLAN	51	1, 5, 14, 38, 67	18
VIII. MONITOR PLAN	20, 75	16, 23, 33, 44, 73	8
IX. INTERPERSONAL RELATIONSHIP			10, 66
X. COMPETENCY	11		
XI. ROLE ACCEPTANCE			21, 36

TABLE 7

GUIDE FOR GENERATING TEST ITEM SCENARIOS[1]

I. *Study the Role Delineation Model to Identify/Think of Critical Incidents*

Note: A critical incident is defined as a set of behaviors that characterize either effective or ineffective performance. Identifying these extremes of a performance dimension in terms of critical incidents helps to understand and to define the performance dimension for measurement purposes.

1.1 Pick an item from the Model. Start with #IA. (Role Area I, Sub-area A) Accept that the role of the PA is limited by supervising physician, legal limitations, etc.

1.2 Think of a situation/incident in which a PA very *effectively* accepted his/her role limitations

or

Think of a situation/incident in which a PA handled his/her role limitations very *ineffectively*.

II. *Write a Test-Item Scenario for this PA Critical Incident*

2.1 *Describe generally* what happened during the incident.

2.2 Define the *conditions* in which this incident unfolds:

2.21 The location/setting (hospital, office, etc.)

2.22 The other health professionals involved

2.23 The type of patient involved (sex, age, socioeconomic status, disposition, clinical condition, etc.)

2.24 The type of health care situation involved (preventive, remedial, rehabilitative, etc.)

2.3 List the *major skills* that the PA needs to handle this situation effectively.

2.4 List some *typical errors*, mis-cues, slip-ups that a PA might succumb to in this situation.

2.5 List some *remedial learning* prescriptions that you would recommend in the case of each error.

---

[1] Prepared by Dr. Ayres D'Costa, College of Education, The Ohio State University.

AAPA
Self-Assessment
Exam

TABLE 8

SAMPLE SCENARIO

Prepared by: ______________
DATE: ______________

Role Delineation Model Code: IV

Critical Incident Description:

Patient is a 11½ month old black male child living in a small community in the northern Midwest. Parents are of lower socioeconomic status. There are two older siblings (ages 1½ and 3 years), and the mother, who is four to five months pregnant, is on welfare; there is no father in the home. The child has been brought to a family practice office for a routine one year old checkup.

The child's weight is 17 pounds, length 27 inches. During the course of the physicial examination you note that his legs are "bowed" with external rotation of knees and internal rotation of the feet. You can elicit full range of motion. The physical exam was otherwise within normal limits. Through a more extensive history, you note that child is on breast milk with the only supplement being orange juice; he eats no solid foods other than baby cereal and crackers. The mother reports that the child does not crawl and makes little attempt to "scoot." Mother reveals she is unhappy about her present pregnancy. She feels hassled and tired and, although is very emotionally caring about her children, feels that her burdens are almost too great to handle.

Conditions:

Disease category - Musculoskeletal
Patient age - Pediatric
Patient sex - Male

Skills Needed:

- Complete nutritional history and social history
- Complete physical exam including hips and extremeties
- Order x-rays of all extremeties and chest

Errors Most Likely:

- Incomplete history
- Limited physical exam
- Inappropriate lab analysis
- Misdiagnosis (i.e., no labs ordered), therefore no treatment

TABLE 9

SAMPLE TEST ITEM BASED ON SCENARIO

AAPA
Self-Assessment
Exam

TEST ITEM

Prepared by:
____
____

Date:________________

Role Delineation Model Code: IV

Item #: 1 Correct Response: C

An 11½ year old black male child is seen in your office for routine physical exam (one year old check). During exam you note external rotation of knees and internal rotation of feet. You can elicit full range of motion, and hips are normal. Otherwise, the physical exam is within normal limits. Your next step should be to:

A. Determine that he has tibial torsion and prescribe orthopedic shoes.

B. Refer to orthopedics for tibial torsion.

C. Obtain radiologic diagnosis to confirm your tentative diagnosis of Ricketts.

D. Refer to supervising physician because you cannot decide what problem exists.

E. Explain to mother that many children have "bowed" legs and that he will grow out of this.

construct item distractors meaningfully. Likes are also established with the remedial learning prescriptions.

Three major types of item *stimuli* were proposed to the item writers: patient conditions/problems scenario, scientific graphic/tabular data/reading passage, and the regular multiple-choice item. These constitute three basic types of stimuli--people/situation encounter, data/report/graphic presentation and the direct verbal question. Each stimulus type has its own peculiar challenges, although the people-type tends to be more unbounded and therefore more complex and challenging. Data and graphic stimuli require specific scientific sophistication and skill, although they can be more straightforward and clearcut. The verbal type of multiple-choice item is the typical examination test item where terminology is important.

Various items styles are associated with each of the above major types of item stimuli, such as classification, relationship or variation analysis, trend/sequence analysis, true-false, five choice completion, five or four choice association, excluded term, quantitative comparison, and multiple completion (K-type). These and other styles were illustrated in the *Guide for Item Writers* thereby suggesting ways to assess different types of cognitive skills. Special efforts were also made to identify multiple-choice strategies to get at interpersonal skills and professional attitudes, e.g., by the use of situations, dilemmas, and best answer items.

The objective of item sampling was not to represent all skills but just those essential behaviors at the terminal level (principle of subsumption). Thus the unnecessary testing of intermediate behaviors was to be avoided in favor of significant generalizable skills with transfer value. Yet the intent of the test was diagnosis and for his reason the test items could not be extremely difficult or representative of above average/excellent performance. The test must represent all entry-level generalist skills in order to represent minimum competence of the PA. Finally, the items must be stratified so as to represent the domain of interest, and random within each stratum in order to be replicable.

Emphasis must also be placed on the proper development of useful response options. It was expected that the typical errors identified for each scenario would lead to the construction of appropriate options. Some of the more common error patterns are: not utilizing all the data provided in the scenario; misinterpreting a technical term; sex-related bias; missing a significant cue; making computational errors including transposing numbers or misplacing the decimal point; using affect-based problem-solving rather than a methodical, logical approach; and inability to handle scientific data correctly.

The documentation of the correct response to an item must be of concern to the test developer. Attempts must be made to validate the correct answer by reference to a standard text, as well as through the process of peer review. Items written by one group of items-writers were reviewed by another group of item-writers. All test items were critiqued in the two try-outs by PAs and by the members of the two Committees. Such critiques point out difficult and esoteric words that creep into items depending on the background and experience of the author. Items that have obnoxious terms or socially undesirable ideas must be modified. Finally there is need to edit items for format awkwardness or inconsistencies, for spelling and grammatical errors and for technical inaccuracies or omissions.

In the case of criterion-referenced tests, there are two somewhat unique item reviews that have to occur, typically by an impartial group of experts: first, a review of the assignment of the item to the specific scale(s) in the specifications matrix. This is a matter of content and construct validity and is critical to the generalization expected in the score interpretation process. Second, a review of the options in each item to identify the correct response option and to identify those options that would be quickly rejected by a minimally competent PA. This latter process is part of the *Nedelsky Technique* (1954) designed to compute an absolute minimum competency score. Nedelsky believed that a group of judges could make such decisions reasonably consistently and thus come up with a dependable minimum competency score. If this is the case, he reasoned that the item is likely to have significant theoretical meaning and the error options then also become educationally significant.

These logical deductions by Nedelsky are pertinent to the construction of a self-assessment examination. It is therefore hoped that the scores derived from the response data will substantiate the true proficiency level of a PA and identify the prevailing error patterns among persons taking the examination.

The item generation process in this project has been very hectic and dependent upon physician assistants most of whom did not have prior experience in test development. Yet the output of some 425 test items, of which 315 were considered reasonably worthwhile to include in the Second Try-Out, is gratifying. Each item is being "banked" in an Item File so that a record of its development is maintained. A sample "item" from the Item Bank is depicted in Figure 5. A file of comments and suggested revisions to items is also being maintained. Wherever appropriate, revisions are being recorded in the Item File.

Each test item is identified by an Item Documentation Card on which are recorded the seven classification criteria for that item as indicated in Table 4. Additionally, this computer card indicates the numbers of the two sets of Scales to which the item has been assigned, the correct response key, its location in the two Try-Out Tests, and any significant recommendations for its future revision/deletion. A sample listing of these cards is presented in Figure 6. The Cards will eventually include the minimum competency score as derived from the Nedelsky Technique.

It is possible to derive a Scoring Key for any scale or for the total test with the help of these Item Documentation Cards and a simple computer program. It is planned to use these Cards as a simple Item Retrieval System so that items of any desired characteristics can be selected using an IBM Sorter. The cards then direct one to the Item File from which a hard copy can be Xeroxed. Obviously this system is not exotic, but we believe it is reasonably flexible and it is inexpensive to maintain.

F. Test Item Revision

Two try-outs have been conducted of the test-items generated for the self-assessment examination. A standard item-analysis program was utilized to generate information about the quality of test-items

FIGURE 5

SAMPLE TEST ITEM FROM ITEM BANK

AAPA
Self-Assessment Exam

Author: ____________
Date: ____________

Role Area: 5
Skill Level: 2
Scale #: 9

Patient Age: 3
MCV Disease: 8
Med. Specialty: 1

Body System: 8
Med. Intervention: 1
Scale #: 4

Text/Reference: Harvey,
P. 611

Items 47, 48

A 22 year old white male college student comes to your office complaining of anorexia and fatigue of recent onset. He has been cramming for finals and jogging to obtain relief from tension. In addition, he has developed wandering joint pains. Further history reveals that he is a homosexual and has been treated for gonorrhea in the past. He denies recent sore throat, urethral discharge and changes in urine or stool color. He also feels that he is sick and tired of smoking and would like some help with that in addition to being treated for his current problem.

= 47.† Your initial differential diagnosis includes all of the following
(127) <u>except</u>:

A. preicteric hepatitis
B. infectious mononucleosis
C. asymptomatic rectal gonorrhea
D. disseminated gonococcemia
E. depression

```
47  TOTAL CORRECT   REL DIFF              1     2     3     4     5     BLNK
   N=   5 PCT=20.0     .800  UPPER        1     1     4     1***  0     0
   CORR PHI=-.482 (SIG= .20)   (PCT)      (14)  (14)  (57)  (14)  ( 0)  ( 0)
   RPBIS=-.443 (ITEM-TOTAL)  LOWER        1     0     1     3***  1     1
   DISCRIMINATION INDICES      (PCT)      (14)  ( 0)  (14)  (43)  (14)  (14)
      OBTAINED D= -28.6      TOTAL        3     3     9     5***  4     1
                               (PCT)      (12)  (12)  (36)  (20)  (16)  ( 4)
```

Items 66-67

A 22 year old white male college student comes to your office complaining of anorexia and fatigue of recent onset. He has been cramming for finals and jogging to obtain relief from tension. Further history reveals that he is a homosexual and has been treated for gonorrhea in the past. He denies recent sore throat, urethral discharge, and changes in urine or stool color. He also feels that he is sick and tired of smoking and would like some help with that in addition to being treated for his current problem.

B47 66. Your <u>initial</u> differential diagnosis includes all of the following <u>except</u>:

A. preicteric hepatitis
B. infectious mononucleosis
C. asymptomatic rectal gonorrhea
D. psychosomatic symptomatology
E. depression

```
66  TOTAL CORRECT   REL DIFF              1     2     3     4     5     BLNK
   N=  28 PCT=25.9     .741  UPPER        1     2     12    10*** 4     0
   CORR PHI= .249 (SIG= .10)   (PCT)      ( 3)  ( 7)  (41)  (34)  (14)  ( 0)
   RPBIS= .100 (ITEM-TOTAL)  LOWER        2     1     16    6***  2     3
   DISCRIMINATION INDICES      (PCT)      ( 7)  ( 3)  (53)  (20)  ( 7)  (10)
      OBTAINED D= 14.5       TOTAL        8     5     54    28*** 10    3
                               (PCT)      ( 7)  ( 5)  (50)  (26)  ( 9)  ( 3)
```

FIGURE 6

TEST ITEM DOCUMENTATION CARD LISTING

Sec#: 2nd try out	Item#: 2nd try out	Sec#: 1st try out	Item#: 1st try out	Correct response	Revision Status	Skill Level	PA Role Area	Patient Age	Med. Intervention	MCV Disease	Body System	Med. Specialty	System Scale	Role Scale
1	127	3	38	A	4	1	4	3	4	12	9	4	17	4
1	128	3	39	A	1	1	6	3	4	12	9	3	17	14
1	129	2	4	[illegible]		2	4	1	2	1	6	5		3
1	130	2	5	D	1	2	6	1	1	2	6	3	1	13
1	131	2	6	E	1	1	4	1	4	1	6	5	13	3
1	132	2	7	C	3	1	5	1	4	16	6	5	13	7
1	133	2	8	D		3	7	1	4	1	6	5	13	13
1	134	1	65	D		2	4	3	1	1	12	1	12	5
1	135	1	66	C		3	2	3	1	1	12	1	12	2
1	136	1	24	D		1	5	2	1	12	9	4	5	10
1	137	1	25	A	1	2	6	2	1	12	9	4	5	14
1	138	1	5	B		2	7	3	1	7	1	2	2	
1	139	1	6	C		2	8	3	1	7	1	2	2	
1	140	5	11	C	1	1	6	3	1	5	6	10	1	13
1	141	5	12	A		1	4	3	4	5	6	1	13	3
1	142	1	52	E	3	2	4	3	1	12	9	4	5	4
1	143	1	56	D	3	2	5	3	1	12	9	4	5	9
1	144	1	54	D	3	2	4	3	1	12	9	4	5	4
1	145	1	53	A	3	2	4	3	1	8	9	4	5	4
1	146	1	55	E		1	8	3	1	12	10	4	11	15
1	147	1	57	D		1	7	3	1	12	9	10	5	14
1	148	1	67	E		2	6	1	1	15	4	3	9	13
1	149	1	54	C		2	5	3	1	15	5	3	3	7
1	150	5	51	C		2	7	3	3	16	5	2	25	13
1	151	5	52	D		2	7	3	3	16	1	2		
1	152	2	15	C		2	4	3	1	16	11	2	6	5
1	153	2	14	E		2	5	3	1	16	11	2	6	10
1	154	5	8	D	1	2	6	4	1	15	6	1	1	13
1	155	2	62	B		1	5	3	4	10	4	1	21	7
1	156	2	2	B		1	4	1	2	1	6	5		3
1	157	5	2	C		2	5	2	3	16	5	2	26	7
1	158	5	4	C	1	2	6	3	4	15	11	1	18	15
1	159	4	19	A		2	5	3	4	16	1	7	14	
1	160	4	20	E		1	5	3	4	16	1	7	14	
2	1	6	42	B		2	5	4	1	15	6	1	1	7
2	2	6	40	D		2	8	3	1	3	6	1	1	13
2	3	6	41	B		2	8	3	1	3	6	1	1	13
2	4	6	8	C		2	6	3	4	7	1	1	14	11
2	5	6	37	B		2	5	3	1	16	7	1	16	8
2	6	6	7	C		2	5	4	1	5	6	1	1	7
2	7	6	32	D		2	5	1	3	16	12	2		16
2	8	6	1	D		2	5	2	1	8	10	4	11	
2	9	1	70	E	5	2	4	3	1	13	6	1	1	3
2	10	1	71	C		2	5	4	1	16	4	3	9	7

and facilitate their review and revision. The Second Try-Out was based upon revisions done on the basis of item analysis of the First Try-Out. The item-analysis data from the Second Try-Out has only been examined cursorily in order to report item quality indices in this Report. We plan to utilize this item-analysis data to revise the items in the Second Try-Out.

A standard Item-Analysis Package was utilized and provided the following types of information:

i) Test Score distribution, including raw score, frequency distribution, cumulative frequency, percentile rank, and standard scores;

ii) Summary Statistics, including mean, median, mode, standard deviation, skewness and kurtosis;

iii) Item analysis, including distribution of responses of upper and lower 27 percent, difficulty and discrimination indices; and

iv) Test quality indices, including reliability, standard error of measurement, distribution of item difficulty and discrimination indices for entire test.

The following discussion is presented in order to explain how the item analysis information was utilized in the development of the self-assessment examination.

It is good to see the test score distribution to get a feeling for the range and clustering of test scores. The measures of central tendency, dispersion, skewness, and kurtosis are also important to decide whether there is a preponderance of masters (hopefully) or nonmasters in the profession. Obviously this assumes that a minimum competency score has been decided upon and can be superimposed on this data.

The item analysis strategy of comparing the upper 27 percent with the lower 27 percent is also valid because it provides a comparison of extreme non-masters and masters. Items with negative discrimination should clearly be avoided. However, there is no reason to select only test items with high discrimination. It must be recognized that in a mastery-nonmastery testing situation, the items are not chosen for their power of separating individuals, but rather because they serve as representatives of critically important responsibilities of the profession. Items must therefore be selected from a narrower range of discrimination, but they must discriminate between masters and non-masters.

One way of implementing this using the regular item-analysis is to compare the lower 27 percent with the remaining 73 percent. Proportionately more of the lower group should choose the wrong options (distractors) than the remaining group. Note that this assumes that in a typical profession, about 70 to 80 percent should be reasonably competent persons, unless there is something seriously wrong with its certification and training process.

Item analysis also provides valuable information about the power of the distractors. Are they serving their function? The fact that certain options are not being selected may indicate that these particular errors/weaknesses have been well-attended to in previous training. Such options should not be excluded. The critical criterion for retaining response options must be relevance to professional skills and pitfalls.

Next, there is the question about the difficulty index or its converse, the number of persons who get the item right. Theoretically, the items should be selected with reference to professional competency relevance, rather than difficulty. It is expected that most normal professions would find a large percentage, say 70 or even 80 percent of their membership competent. Therefore, the items in such a test would appear easy.

Thus, we would argue that the three Ds of test construction--discrimination, distractors, difficulty--are also of importance, albeit in a very different way, in criterion-referenced test development. As with those who ignore the lessons of history, those who choose to ignore such data would stand condemned by them.

In this project, the item developers were provided a simplified summary report, based on the item analysis data for each test item. The three Ds were presented by using certain codes. Also summarized for each item was a code which indicated if the item was questioned in terms of its relevance to the PA profession by those PAs who took the First Try-Out examination. It will be remembered that each item was separately rated for relevance on a five-point scale. These ratings were summarized across raters.

Finally, we have the overall test quality indices, especially reliability and standard error of measurement. Standard texts on criterion-referenced testing warn against the use of the traditional methods of computing reliability, such as the correlation coefficient. Suspect also are measures such as Kuder-Richardson Formula 20 and Hoyt's Index. The main reason for this warning is because of the deliberate reduction in variance that occurs in criterion-referenced tests. Other indices are therefore proposed such as "Kappa" (Cohen, 1960) and "coefficient of agreement" (Subkoviak, 1976). Hambleton and Eignor (1979) suggest the Subkoviak approach when a test is only administered once. Accordingly, a computer program was written to compute this coefficient for the total test, for each of the 17 Role scales, and for each of the 28 System scales of the Self-Assessment Examination.

The "coefficient of agreement" was originally defined as the probability that each individual in a group will be consistently classified as a "master" or "non-master" on two successive parallel tests. However, like the Kuder-Richardson Formula 20, it is possible to estimate this index (coefficient of agreement) from a single administration of the test by using the assumption that all items are equally difficult or reasonably so. The decision to classify as a "master" is based on achieving the minimum competency score identified through the Nedelsky Technique by PA experts.

The coefficient of agreement is similar to "Kappa" and must be interpreted like any probability value with a range from 0 to 1. "Kappa" is an index of reliability appropriate for criterion-referenced tests.

The standard error of measurement is a critical index to present in any test development effort. Hambleton and Eignor (1979) explain that this index is valid for criterion-referenced-testing as well.

G. *Test Development Statistics*

The data gathered as a result of the various analyses conducted in the development of this test are far too voluminous to present in this Report. Instead the following selected summary Tables will be presented as Exhibits without discussion:

i) Frequency distribution of Total Test Scores (Exhibit C);

ii) Means, standard deviations, and Minimum Competency Score for the 17 Role Scales (Exhibit D);

iii) Means, standard deviations, and Minimum Competency Score for the 28 System Scales (Exhibit E);

iv) Distributions of Item Difficulty and Discrimination indices for Total Test (Exhibit F);

v) Reliability, Coefficient of Agreement, and SEM for Total Test and for the 17 Role Scales (Exhibit G)

vi) Reliability, Coefficient of Agreement, and SEM for the 28 System Scales (Exhibit H)

vii) Assignment of Second Try-Out Test Items to original Test Specifications Matrix (Exhibit I)

H. Test Interpretation

"The most obvious benefit of self-assessment", wrote Hess and Morreau (1976), "is that with minimum personal consequences physicians can readily ascertain what they know (or do not know) in given areas--areas where their future decision-making may have profound consequences." This self-assessment examination for physician assistants requires about six hours of personal time investment, but the payoff could consist of good information in several dozen areas (scales)--information that could assure competent professional care. To be useful, information must be reliable, valid, complete, relevant and usable.

The matter of reliability has already been dealt with. It concerns consistency in making judgments from one occasion to another. Judgments must agree with one another if they are to be reliable. The coefficient of agreement provides data on this matter.

We are concerned here with the proper utilization of the results of this examination and the issue of validity becomes central. Typically, criterion-referenced testing has been limited to content validity based upon expert judgment. The process of scale-development takes us beyond this to construct validity where our interest lies in the underlying scale area. The process to-date has utilized largely the judgment of expert Committees to assure us that the items do indeed represent and will therefore predict the scale area. The discrimination index computed in the item analysis provides some validity insights as well. If a test item is a valid measure of a construct, then persons who do well on the overall scale should get that particular item correct as well. Although this reasoning is somewhat circular it is nevertheless useful.

We do not yet have data to report that would assure us of the predictive power of the test. Our interest is largely in short-term prediction at this time because the main purpose is to motivate the professional to undertake the continuing education he needs or wishes. To say that persons who do well in the self-assessment exam will also perform well on the job seems like a tall order. There are too many other necessary conditions that must be met before undertaking such predictions.

The concept of validity conjures up notions among researchers of being able to draw justified inferences (internal validity) and to generalize beyond the present circumstances (external validity). Translating these notions to this exam, one asks: Is success in the exam attributable entirely to possession of the appropriate professional knowledge and skills? We know that this is not necessarily true. Lack of test-wiseness, mental stress, preoccupation with other matters, and physical status can all affect how one fares in a test. Fortunately, a self-assessment exam reduces many of these usual test performance problems because the exam can be taken at one's leisure and without time pressures. However, the problems of generalizability remain unless the test item and the expected performance standards are fair to all individuals in the profession. Problems of generalizability can occur because of differences in practice characteristics, or because of personal interests and motivations. Because of the recognition of these problems, this particular self-assessment has planned a three-way approach to deriving continuing education prescriptions, namely, practice (P) characteristics, personal interests or needs (N), and examination-derived indices (E).

This three-way analysis ensures relevance and completeness to the self-assessment. Too often such assessments are limited to tests and they fail to recognize the importance of the other critical forces that impinge on the decision to seek continuing education. Perhaps a fourth factor in the AAPA program lies in the built-in awarding of credit, a positive reinforcement that will strengthen the three-way process further.

We are then left with the matter of usability of the information derived from the self-assessment. This project has developed a computer-generated four-page report which provides answers to the following types of questions (See Exhibit J):

1. How did I do on the test as a whole? (Total Test Score) How do I compare with other PAs? (Mean and Standard Deviation for a national sample of PAs, Page 1);

2. Which items did I get wrong? What was the right answer? (Print-out of the individual's responses with correct answers keyed to each response, Page 1);

3. How did I fare in specific areas pertaining to my role as a PA and where am I expected to perform satisfactorily? (Print-out of scores on 17 Role scales compared with maximum possible score and minimum competency score, Page 2);

4. How did I fare in terms of my knowledge and skill relative to the major Body Systems? (Print-out of scores on 28 System scales compared with maximum possible score and minimum competency score, Page 3); and

5. How do my exam scores (E) compare with my personal need scores (N) and with the clinical practice (P) I am engaged in? (Graphic display of the P, N, E converted scores for each of the 28 Body Systems Scales, Page 4).

Accompanying this four-page report is an *Interpretive Leaflet* which helps the individual PA understand the report and utilize it for his self-assessment (See Exhibit K). The leaflet follows the sequence of the four-page report and explains how each piece of data might be utilized by the PA. Attempts were made to avoid using technical terms except where absolutely necessary. Any attempt to talk down to a PA had to be avoided as well. Accordingly, some technical terms, such as reliability and standard error of measurement, were retained. These are critical to the proper interpretation of test scores and have become part of common scientific language.

Several technical questions will probably arise at this time:

1. Is it technically correct to report raw scores? Are they valid measures of the constructs they represent?;

2. What about the reporting of Mean and Standard Deviation of the PA national sample? Isn't this going back to norm-referenced testing? How representative was the sample?;

3. How was the minimum competency score derived? How do I interpret my own score relative to this standard? How reliable and valid is this comparison?; and

4. How were the graphic indicators of P, N, and E derived? Is this comparison appropriate for criterion-referenced testing?

The use of raw scores in reporting test results has become quite commonplace in recent years. Interest tests like the Ohio Vocational Interest Survey (D'Costa, 1969), the Strong-Campbell Interest Inventory (Campbell, 1972), and others have preferred to use the raw score instead of a group-referenced score like the standard score because it is recognized that the major utility in comparisons lies within the individual's own system of needs and preferences. Often the individual wishes to compare his strength in one area with another area without reference to his relative standing in his peer group. A dermatologist does not care to know that his knowledge of skin diseases is superior among all physicians. He does care for comparisons with other dermatologists, but more importantly he is interested in knowing his strengths and weaknesses relative to himself alone in order to determine what thrust his practice should take.

The matter of validity of raw scores is then based upon their relevance for the type of decisions that must be made. This is particularly true for criterion-referenced tests since the major part of the interpretation of such tests is located in the criteria or constructs being measured. Given the fidelity to criteria inherent in the test development process, such interpretation therefore becomes technically appropriate and defensible. However, it must be noted that the main interest is not in the precise differences between two scores but in their relative distance from their minimum competency scores which serve as their points of reference. Compare this with the use of group mean scores as the points of reference, and it becomes evident why the raw score coupled with its minimum competency score is appropriate for self-assessment.

The reporting of the mean and standard deviation of a national sample is essentially to satisfy individual curiosity and make the self-assessment more interesting and meaningful to some individuals who must use this type of indirect peer pressure to motivate themselves. This is a norm-referenced technique and its validity depends upon the representativeness of the national sample of PAs.

The norm-group in the Second Try-Out is a sample of 108 PAs who voluntarily responded to the self-assessment exam as of July 1, 1979, out of some 274 PAs who were selected on the basis of a stratified random sample of the AAPA membership. Effort was made to represent various geographic locations and practice characteristics. Excluded from the sample were persons who were involved in the test-development process. In terms of number of years in practice, the responders were distributed as follows:

Less than 1 year	20 percent,
One year or little more	4 percent,
About two years	18 percent,
About three or four years	25 percent, and
More than four years	34 percent.

In terms of type of patient care provided,the responders were distributed as follows:

Medicine (family, general, internal)	78 percent,
Surgery	20 percent,
Pediatrics	2 percent, and
Obstetrics/Gynecology	1 percent.

The above data has been rounded off to the nearest integer and hence the totals add to 101. No data is currently available by which to judge the representativeness of this sample with respect to the PA population. Informal opinions indicate that the two distributions are not surprising. Until a national profile of PAs is developed and becomes available it is difficult to make such comparisons. However, the sample does include PAs with a wide range of practice experience. Also, it is encouraging to note the large percentage of generalists in the sample.

The computation of the Minimum Competency Score is based upon the Nedelsky Technique. In essence it amounts to recognizing that in a five-option test, a minimally competent professional must be capable of rejecting some of these options because he immediately recognizes the erroneous thinking in them. If three options are expected to be rejected, then only the remaining two options serve as legitimate distractors and the expected score on this item is therefore 0.5. Using this approach with each item, it then becomes possible to compute the expected score on the total test for a minimally competent PA. The term "minimally competent PA" is intended to indicate basic or expected level of skills for an entry-level generalist PA.

The judgments are made by experts who in this case were the members of the Test Specifications Committee; two are physicians, and three are PA educators or PAs. Each rating was made independently and later discussed at the Committee meeting in June.

Technically, a standard such as a minimum competency score is treated like an absolute cut-off score. You have either attained it or you have not. In self-assessment, there is no need for such absolutism and its concommitant hazards. For every test score, measurement experts point to a standard error of measurement. It is recognized that one's true score may lie within a range of two standard errors in about two-thirds of the cases. Thus if one's score is 200 and the standard error is 7, the true score may lie within 193 and 207. Wider ranges are prescribed if greater confidence, or a lesser margin of error, is required.

It is therefore recommended that a raw score distance from a competency standard be interpreted in terms of the standard error of measurement. If your raw score is 200 and the competency standard is 205, while the standard of error of measurement is 7, it should be recognized that here the discrepancy is not large enough to cause worry about one's competence. As such a discrepancy approaches 3 or more multiples of the standard error of measurement, serious concern should occur.

In the case of the results of particular examination, it must be recognized this exam is still in its Try-Out stage of development and the current process of revising items and validating the scales must be moved forward before greater confidence can be placed in the scores. Likewise the minimum competency scores might need revision based upon reaction from PAs-in-practice to them.

The reliability of discrepancy scores has been a thorny problem in measurement because of the large error statisticians associate with them. Where reliability is weak, it is difficult to get good validity as well. Indeed reliability is a prerequisite for validity to occur. The discrepancy data provided in the report is therefore not to be interpreted literally but in context. The questions to ask are: Is this really true of myself? Is there a difference in the way I think through such problems compared to my peers? Is this important to me, as a professional, in my practice? Where can I get more information and assistance?

At the current time, the AAPA might be able to identify and offer a few CME programs to PAs interested in following up their self-assessment reports. The ultimate objective is to link CME module recommendations to the self-assessment so that effective follow-up is possible. It is hoped that a modular system of learning packages can be developed so as to have proper linkages with the *Role Delineation* .

The graphic indicators of P, N, and E scores are based on standard scores which are technically known as "linear stanines". Stanines were popularized by testing programs in World War II days. The word "stanine" is derived from the *stan*dard *nine* points of reference utilized in this technique. In norm-referenced testing, stanines are constructed so as to be interpretable in terms of the normal curve. Linear stanines do not assume a normal curve and are based on a simple linear transformation of scores so that the new mean and standard deviation are always 5 and 0.5 respectively.

By converting all the scores used in the graphic display of P, N, and E into the stanine system we obtain comparability both within the P, N, and E indicators for one Body System Scale and between Body System Scales. All the indicators can be compared with one another because they have been transformed into this stanine system. This capability to compare indicators is the crux of this P, N, and E report. The intent is to allow the individual to make comparisons and check them out in terms of his own internal beliefs about himself. Internal comparisons serve as stimuli rather than indictments about oneself and become the essential core of self-assessment.

The appropriateness of group or norm-referenced information in a criterion-referenced test is not a technical concern among experts like Popham (1976) who noted that normative information often provides additional insights into what should constitute an acceptable level of performance. The power of a criterion-referenced test lies inherently in its ability to describe what the individual can do, and the addition of normative data adds to these insights. On the other hand, norm-referenced tests by themselves cannot provide such individual descriptions and are therefore weak as diagnostic tools.

What kind of comparisons can an individual make with P, N, E indicators in the computer-generated report and how does one go about deriving them? (See the *Interpretive Leaflet* for specific suggestions for comparisons). For these insights it must be recognized that the indicators are not to be used as precise measures. Measurement errors are of concern here too, although the scale is much reduced. The indicators are drawn proportionately in nine different lengths to represent the nine stanines. In norm-referenced interpretations, 4, 5, 6 are considered average, while 1, 2, 3 are below, and 7, 8, 9 are above average.

Technically speaking, linear stanines are subject to aberrant values when the distribution of scores is markedly skewed positively or negatively. The computer program was designed so that computed stanine values in such situations did not go below 1 or above 9. This approximation was introduced for the sake of reporting convenience.

I. A Review of Technical Limitations and Deficiencies

The self-assessment examination for PAs, as currently developed by the AAPA under this contract, has several limitations which need to be acknowledged:

i) The quality of the test item styles needs to be improved. This criticism has to do with the limited test item styles utilized by the item writers. More items need to be written involving graphs, charts, and scientific tables. There are too many K-type (multiple completion) test items and many of these do not take advantage of this particular item format.

ii) The process of item revision needs to be continued much further. The revisions to-date are primarily based on one try-out and expert opinions. The item analysis from the second try-out must be utilized and the process of revision continued until the item statistics, particularly the few negative and several very low discriminations that still remain are removed.

iii) The quality of item options (distractors) must be improved so that an effective system of error patterns analysis can be developed.

iv) The scenarios developed must be reviewed for criticality and representativeness of PA performance. Only a few beginning steps have been implemented relative to this interesting technique.

v) There are not enough items identified for the item bank. Several scales have less than 10 items per scale. This limits the reliability, relevance, and the representativeness of the scales that have been proposed for the self-assessment model.

vi) The scale development process has been entirely judgmental. No empirical analysis has yet been undertaken to ensure the proper allocation of items to scales or to ensure the homogeneity of the scales. Factor analytic approaches are available to generate scales for such multifactor test batteries. This criticism also applies to the domain definition process. The constructs pertaining to the scales are judgmental and lack empirical validity data at this time. Some scales may need to be deleted and others added as the entry-level generalist PA role gets better defined.

vii) More developmental studies are needed before the diagnostic qualities of this examination can be considered satisfactory in terms of professional measurement standards. In particular, deficiencies are noted in terms of reliability and validity data for the various scales. Many of the scales appear to have very weak reliabilities at this time. Several studies that can be done with the current data have not yet been done.

viii) Although the content for this Report was in part derived from the 10-point checklist provided by Hambleton and Eignor (1979) for rating criterion-referenced tests, it must be recognized that the weaknesses, as noted above, limit the quality of this Report as a Technical Manual for the Self-Assessment Examination.

IV. USING THE SELF-ASSESSMENT EXAMINATION TO DEVELOP A PILOT CME SYSTEM FOR PAs

A. *Rationale*

The AAPA subscribes to a life-long system of continuing education for its membership in order to assure the prestige of its profession and to ensure its role in providing quality health care to society.

The development of a national continuing education system for PAs entails several issues, of which the following appear to be the most critical to the PA profession at this time:

1. What kinds of CME needs do PAs have? Is there some pattern to these needs in relation to practice length (time elapsed since certification), type of practice (especially supervising physician speciality), geographic setting, practice location?;
2. How are these CME needs related to performance needs? Will the CME proposed result in the desired quality of health care?;
3. Are PAs aware of their CME needs? What kinds of CME do PAs normally seek? How much?;
4. How do PAs obtain their CME at this time? What approaches seem popular, valued, disliked? What kinds of CME programs are currently available to PAs? How good are they in terms of meeting the needs of the profession?; and
5. How can a national CME system be developed so that a life-long (graduation gown to grave) competency assurance program is available and utilized by most PAs?

Logistically, the identification of CME needs of the PA profession can be effected by using self-assessment examination data. However, this will require a major national effort because it goes way beyond the usual professional survey in depth, although somewhat similar in extent. It is imperative that the approach to PAs be made in non-threatening terms and with sufficient utility offered to make their participation possible and worthwhile. With participation time and data-gathering costs becoming increasingly ominour, it is necessary to come up with innovative approaches to data-gathering, which will fit into the professional style and schedules of PAs and yet satisfy the needs of statistical inference and generalization.

Theoretically, a national profile of PAs must be valid. The domain on which the profile is generated must be relevant and acceptable to the profession. Given a new and developing profession this task is not easy. Given the role of the supervising physician in the role of all PAs, it becomes necessary to recognize this fact in the process of domain definition but without diminishing the stature of this group as a profession. Ths changing pattern of health care services in this country further complicates this task. With national health insurance looming not too far off, the PA is bound to be called upon to modify his role relative to this national health care need. As more of the new type of allied medical professionals are ushered into the health care system, role changes and new responsibilities will occur.

Yet, within this dynamic system an assessment of the quality of the profession appears very much in order. Considerable public investment has gone into the creation of this new professional. Expectations remain high and it is therefore legitimate to embark upon a reasonable effort to provide accountability data to the public.

Professionally, there is nothing more challenging that the opportunity to demonstrate that this young PA profession is conscious of its responsibilities and willing to do whatever is necessary to maintain quality in its ranks. The AAPA has already embarked upon several continuing education programs and would welcome the opportunity to organize the necessary national effort and recruit membership support.

In generating new programs for the profession, however, the AAPA must face the fact that it is responsible to each individual member. The national interest in the quality of the profession must therefore be based upon the natural interest of each member to remain a worthy and useful member of the profession.

B. *Methodology Proposed*

The next step in developing the CME system for PAs should take the form of a national pilot program founded upon the theoretical, logistic, and professional considerations discussed above.

Products that need to be developed in the theoretical domain were "brainstormed" by the Test Specifications Committee at both its meetings. The major need is to relate professional responsibilities to professional performance. The task is to find the linkages between how PAs handle their clinical responsibilities and the quality of their knowledge, skills, and attitudes relative to these same responsibilities. This calls for an indepth analysis of clinical performance along with an indepth diagnosis of knowledge, skills, and attitudes.

Several strategies are available to the AAPA for implementing such an analysis of the causal dimensions of professional performance. Included would be the selection of a few representative PA clinical training programs so that performance assessments can be made by clinical supervisors, and the skills, knowledge, and attitudes assessments can be handled by a revised national self-assessment examination.

Needed for such an analysis are clinical assessment tools and revised forms of the self-assessment examination. It would be imperative to ensure that the domains assessed by the two types of assessments are compatible, and that they in turn relate to the *Role Delineation for Physician Assistants* . The adaptation of the *Role Delineation* so that it might better fit such a CME model was already begun in this current project. The eleven areas of responsibility were reclassified into three comprehensive areas of competence--professional, interpersonal, and clinical. Each competency area includes three or more of the original eleven areas of responsibility. The "clinical" area, however, can also be subdivided by Body Systems. The 17 Role Scales developed in the Self-Assessment Examination are based upon this adaptation of the *Role Delineation* .

The development of this model is far from complete at this time. There is need for further classification so that the skills identified in the *Role Delineation* are better represented and assessed both in the Self-Assessment Examination and in the set of clinical assessment tools that must be assembled.

The implemention of such a project, from a logistics standpoint, calls for a national effort with collaboration of selected PA training programs. It would not be difficult to gather data from PAs-in-training for such a project. However, it would be unwise to limit these "causal analyses" to such groups alone. The need to generalize such findings to practicing professionals requires that the major study be conducted with practitioners. The training programs would only be involved in the "causal analysis " component of the project.

Aside from generating the basic causal model, the project should aim at identifying the major CME needs of PAs and setting up a pilot model for implementing an appropriate CME program for PAs. This calls for three other components to the project: development of self-assessment examinations, obtaining data for a national PA competency profile, and development of CME learning packages.

The national PA profile would essentially amount to a CME needs assessment. It would serve as the basis for emphasis in the development of learning packages. The learning packages would be modularized and geared to the self-assessment examination. The self-assessment examination would also be modularized so as to make it convenient to take, receive feedback, and follow-up by CME. Appropriate linkages would need to be developed so that the causal model is operationalized and thus CME is given the chance to result in better professional performance.

The professional considerations in implementing such a project require that it receive the support of the professional organizations concerned and of their membership. Rather than initiate a massive new effort with all the concommittant hazards and start-up costs, it would be prudent to work into existing professional CME systems and available CME mechanisms without getting overly bogged down in them. This project would need their support but not necessarily their burdens. The two, however, do not always exist separately.

Opportunities, such as national meetings, currently available expertise, interests, and products, should be taken advantage of. It must particularly be recognized that other health professions may have already dealt with some of these problems and that such know-how is transferable at lesser cost of time, money, and people.

C. *Time Schedule*

It is estimated that the schedule for the implementation of this proposed CME pilot system would require about three years with achievements targetted approximately as follows:

Year 1	Develop all needed tools including their try-out. Work with PA training programs to establish methodology. Develop test item bank for self-assessment exam. Survey CME approaches, methods, and offerings.
Year 2	Conduct try-out of the national profile of PAs. Develop strategies. Develop and try-out learning packages. Develop feedback system.
Year 3	Complete national profile. Develop learning packages; modify packages. Use CME feedback system on pilot basis. Evaluate and recommend CME system.

V. FINDINGS AND CONCLUSIONS

The purpose of this component of Contract HRA 231-76-0053 was to develop a criterion-referenced self-assessment examination for physician assistants, using the *Role Delineation* as the basis, so that appropriate learning prescriptions could be developed in order to facilitate the continuing medical education of physician assistants and thereby ensure the quality of health care provided by them.

A 300-item comprehensive examination was proposed for development using two Working Committees consisting of PAs and PA educators. The Test Specifications Committee provided general guidelines for the development of the examination and the Item Writers Committee did the major work of writing and revising the test-items generated by the project. The examination has undergone two try-outs and has been revised each time, but additional revision is planned with the extensive data now available.

The self-assessment examination is based upon the domain defined in the *Role Delineation for PAs*. Two sets of scales have been generated, thereby allowing two approaches to the specification of the domain. One set of scales, Role Scales, is based upon an adaptation of the eleven major responsibilities of the PA. The other set of scales, Body System Scales, is based upon the matrix comprising Body Systems and Medical Intervention types. The 315 items that were administered as part of Form A of the Self-Assessment Examination were assigned with the help of expert judgment to each of the two sets of scales and scores were generated.

The interpretation of the scores obtained on the 17 Role Scales and the 28 Body System Scales is done with the help of minimum competency scores which were determined by expert judgments using the Nedelsky Technique. It must be acknowledged at this point that this effort needs additional data-gathering and development.

Two somewhat innovative approaches were used in the implementation of this project. One has to do with the process of item generation where the critical incidents approach was used. The other relates to the conceptual model for CME using self-assessment examinations. It is the thesis of this model that CME must be based upon a combined analysis of practice (P) requirements, individual felt needs (N), and deficits identified by examination (E) scores. In accordance with this model, a four-page computer-generated reporting system was developed and returned along with an *Interpretive Leaflet* as feedback to PAs who participated in the second try-out of the examination.

The test-development process has been constrained by time but has nevertheless attempted to adhere to the professional standards prescribed by measurement specialists (APA-AERA-NCME Standards). The standards identified for criterion-referenced test development by Hambleton and Eignor (1979) were also recognized in this project. Several limitations have been acknowledged relative to the "quality" of the test items in the current form of the examination. However an overall assessment of the examination must acknowledge not only its future potential but also several current good qualities.

Although this Report must serve also as the Technical Manual for the Examination, limitations must be acknowledged in this respect. Not all of the research which should be contained in

best copy available

such a Manual has been completed. Nor has there been sufficient time and opportunity to assimilate all the data and analyses available to-date in order to provide a good discussion of the tables currently included in the Exhibits. It is hoped that the inquisitive reader will use the appended data and direct comments and inquiries to the AAPA so that the needed technical reports can be developed and added at a future date.

The *Products* generated by this project have been listed elsewhere; however, by way of summary, it needs to be noted that the AAPA now has an Item Bank in the process of development, and a feedback system for PAs taking the self-assessment examination which should serve as a first step towards their continuing education. The suggestions given in this Report for a Pilot CME System are based upon the experiences generated in this project and upon a national perspective of the expected directions and needed next steps for the physician assistant profession.

Finally, the reader might ask: How has this project been *evaluated* by PAs? Anonymous comments were received from PAs who participated in the First Try-Out. Most of these were favorable and indicated that the membership was pleased with this undertaking. Formal data was gathered from the 108 PAs who took the entire three-section self-assessment examination. We were concerned because this was a randomly selected group who had not volunteered for this imposition nor was it possible to assure them any CME credit because approval had not yet been received for such credit at that time. (The self-assessment examination has since been granted six hours of Category I credit, and this good news will be added to the feedback package that will be mailed to each PA in early September). Table 10 presents the evaluative data that has been summarized from Items #59, 60, 61, and 62 of Section 3 of the Self-Assessment Examination.

The following summary conclusions and recommendations can be extracted from those data:

1. The self-assessment examination required about 6 hours of work from the PAs who participated in the Second Try-Out;

2. The validity of the exam (using weights of 1, 2, 3, 4, 5 for the 5 ratings and averaging them) was rated at between Good and Satisfactory (3.5);

3. Only 7 out of the 102 PAs who responded to this item responded with a "maybe" to the question whether AAPA should continue its efforts developing such self-assessment exams. About 75 percent said "Yes, very useful" and 20 percent said "Yes, somewhat useful". (It must be recognized that these ratings occurred without any feedback being received, and after the completion of a somewhat arduous task); and

4. About 78 percent requested as much feedback as the AAPA could afford to send them. This question also had two interesting response options indicating "specialty" comparisons versus all PAs' comparisons. It is noteworthy that 21 want additional comparisons relative to their own specialty whereas only two want comparisons with all PAs only.

TABLE 10

EVALUATIVE DATA

374. **Approximately how many hours did this exam require of you?**

35	A. Less than 6
43	B. About 6, but less than 7
16	C. About 7, but less than 8
6	D. About 8, but less than 9
3	E. About 9 hours, or more
5	Blank

375. **Recognizing that this Form A is still somewhat new, how valid (in terms of content, quality of questions, and type of exam) would you consider this exam to be as a self-assessment device for PAs?**

1	A. Very poor
15	B. Weak
40	C. Satisfactory
33	D. Good
14	E. Very good
5	Blank

376. **Should the AAPA continue its efforts to improve this form and to develop new self-assessment exams like it in the future?**

76	A. Yes, this would be very useful
19	B. Yes, this would be somewhat useful
7	C. Maybe, but I'm not too sure
0	D. No, it is not very useful
0	E. No, it is a real waste of time
6	Blank

377. **How much reporting of results (feedback) would you like to receive regarding your performance on this exam?**

79	A. As much feedback as the AAPA can afford to send me
21	B. Information about my own performance with comparisons to all PAs, and especially PAs in my own specialty
2	C. Information about my own performance with comparisons to all PAs only
1	D. Just my own raw scores and subscores
0	E. I don't care to receive any feedback
5	Blank

REFERENCES

Bloom, B.S., ed. Taxonomy of Educational Objectives: Cognitive Domain. New York: David McKay, 1956.

Campbell, David. Strong-Campbell Interest Inventory. Minneapolis, Minn.: University of Minnesota, 1972.

Cohen, J. "A Coefficient of Agreement for Nominal Scales." Educational and Psychological Measurement, 1960, 20, 37-46.

D'Costa, Ayres et al. Ohio Vocational Interest Survey. New York: The Psychological Corporation, 1970.

D'Costa, Ayres. The Longitudinal Study of Physicians. Technical Proposal funded by NCHSR, Association of American Medical Colleges, 1975.

Eignor, D.R. and Hambleton, R.K. "Effects of test length and advancement score on several criterion-referenced test reliability and validity indices." Laboratory of Psychometric and Evaluative Research Report No. 86. Amherst, Mass.: School of Education, University of Massachusetts, 1979.

Engel, J.D. "A Comparison of Diagnostic and Certifying Examinations." American Journal of Medical Technology, 43, 5, Dec. 1976, 436-439.

Flanagan, John C. "The Critical Incident Technique." Psychological Bulletin, 51, 4, July 1954, 327-359.

Gagne, R.M. The Conditions of Learning. New York: Holt, Rinehart, 1970.

Hambleton, R.K. and Eignor, D.R. "Competency Test Development, Validation, and Standard Setting." In R. Jager and C. Tittle, eds., Minimum Competency Testing. Berkeley, Calif.: McCutchan, 1979.

Hambleton, R.K. and Eignor, D.R. Criterion-Referenced Test Development and Validation Methods, AERA Training Program Materials, 1979.

Hess, K.M. and Morreau, L.E. "The Expanding Classroom...Why Self-Assessment?" Postgraduate Medicine, 59, 1, Jan. 1976, 203-210.

Krathwohl, D.R., Bloom, B.S. and Masia, B.B. Taxonomy of Educational Objectives: Handbook II: Affective Domain. New York: David McKay, 1964.

Marsland, D.W., Wood, M., and May, F. "Content of Family Practice." Journal of Family Practice, 3, 1976, 37-45.

Milgrom, P., Weinstein, P., Ratener, P., and Morrison, K. "Dentists Self-Evaluations: Relationships to Clinical Performance." Journal of Dental Education, 42, 4, 1978, 180-185.

Millman, J. "Criterion-Referenced Measurement." In W.J. Popham, ed., Evaluation in Education: Current Applications. Berkeley, Calif.: McCutchan, 1974.

Nedelsky, L. "Absolute Grading Standards for Objective Tests." Educational and Psychological Measurement, 14, 3-19, Spring 1954.

Perry, Henry B. "An Analysis of the Professional Performance of Physician's Assistants." Journal of Medical Education, 52, Aug. 1977, 639-647.

Popham, W.J. Criterion-Referenced Measurement. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

Pottinger, P.S. Competence Testing as a Basis for Licensing: Problems and Projects. Paper presented at Conference on Credentialism, Berkeley, Calif., April 1977.

Subkoviak, M. "Estimating reliability from a single administration of a criterion-referenced test." Journal of Educational Measurement, 1976, 13, 265-275.

Wilson, Margaret A. "Basic Principles of Credentialling Health Practitioner." Respiratory Care, 21, 10, Oct. 1976, 954.

EXHIBIT A

UNPRIORITIZED LIST OF PROGRAM AND RESEARCH OBJECTIVES

Program Objectives

1. Develop appropriate baseline for recertification exam.
2. Profile PAs national/CME needs assessment.
3. Provide individualized feedback.
4. Evaluate PA training programs.
5. Check if exam matches clinical practice.
6. Develop learning packages to improve competence.
7. Use log diary to validate test.
8. Check individual's learning style to provide CME.
9. Check if competence improves as a result of learning package.
10. Collect critical incidents and skills and relate to curriculum.
11. Note sensitivities of PA profession to test items.
12. Conduct item analysis.
13. Do key word analysis.
14. Prepare two equivalent forms of the test.

Research Objectives

1. Compare option formats.
2. Compare item styles.
3. Check strengths/weaknesses by geographical location, training program, specialty.
4. Compare self-assessed competence and successful recertification.
5. Develop test specifications matrix.
6. Develop pragmatic research model to research profession.
7. Study causal dimensions of competence.
8. Study relationship between competence areas and tasks frequently done.
9. Develop item pool.
10. Identify core performances.
11. Study longitudinal changes in PA population on all parameters.
12. Prepare profile of test components (content).
13. Provide referenced feedback.

EXHIBIT B

THE ELEVEN MAJOR RESPONSIBILITIES OF THE PA ROLE DELINEATION

I. RECOGNIZE INTERDEPENDENT RELATIONSHIP WITH SUPERVISING PHYSICIAN

   A. Accept that the role of physician assistant is limited
   B. Resist compromises in the practice of medicine when conflicting with professional ethics
   C. Express professional opinion on matters of patient care, even if different from supervising physician's opinion
   D. Express limitations of the role when necessary

II. DEMONSTRATE PROFESSIONAL BEHAVIOR

   A. Possess attributes of empathy, objectivity, tolerance, confidence
   B. Demonstrate professional attributes in actions

III. PROMOTE PREVENTIVE HEALTH CARE

   A. Educate patient and family concerning health care measures
   B. Perform screening examinations
   C. Provide sex education
   D. Provide counseling to patient and family
   E. Provide resources for patient education

IV. ESTABLISH HEALTH STATUS DATA BASE

   A. Modify data gathering process as necessary
   B. Elicit pertinent medical and psycho-social history
   C. Perform physical examination as pertinent
   D. Establish preliminary diagnosis of common problems
   E. Obtain information from screening and diagnostic tests by ordering and performing tests and obtaining specimens
   F. Record and transmit findings from history and physical examination
   G. Inform physician of tentative problem list

V. ANALYZE DATA BASE

   A. Differentiate between normal and abnormal (including variations of normal) information contained in the data base
   B. Interpret raw data from screening and diagnostic tests
   C. Interpret written report of screening and diagnostic tests
   D. Validate preliminary diagnosis of common problems
   E. Develop diagnostic impressions from information contained in the data base
   F. Establish working diagnosis of common problems
   G. Confer with supervising physician according to practice's guidelines

EXHIBIT B CONTINUED

VI. FORMULATE HEALTH MANAGEMENT PLAN

A. Resolve deficiencies defined by data base
B. Prioritize problems to be managed
C. Devise plan to coordinate multiple treatment modalities
D. Select therapeutic measures
E. Select supportive services to be involved in patient care
F. Describe parameters of patient education relating to immediate problems, then others
G. Formulate a management plan for common problems

VII. IMPLEMENT HEALTH MANAGEMENT PLAN

A. Educate patients and family
B. Contact supportive services to be involved in patient care
C. Provide information pertinent to consultation/referral
D. Provide treatment of common problems
E. Refer patients as necessary for treatment of common problems
F. Initiate medical therapies/procedures

VIII. MONITOR HEALTH MANAGEMENT PLAN

A. Assess degree of patient compliance
B. Assess progress toward desired result
C. Determine economic impact of management plan
D. Determine impact of community resources
E. Recognize undesirable effects of treatment plan
F. Redirect patient efforts based upon results of treatment plan

IX. ESTABLISH EFFECTIVE INTERPERSONAL RELATIONSHIPS WITH PATIENTS, PROFESSIONALS, AND OTHERS

A. Adapt suitable interviewing style
B. Accept personal, cultural, and professional factors affecting health
C. Assist patient/family in handling/expressing feelings
D. Recognize changes in patient's psychological state
E. Maintain relationship with referred patients
F. Demonstrate concern for patient's privacy, modesty, anxieties during the examination
G. Transmit and record information

X. MAINTAIN COMPETENCY

A. Engage in periodic review of professional skills (self-assessment, etc.)
B. Devise and maintain program of formal and informal CME based upon recognized needs

EXHIBIT B CONTINUED

X. MAINTAIN COMPETENCY (continued)

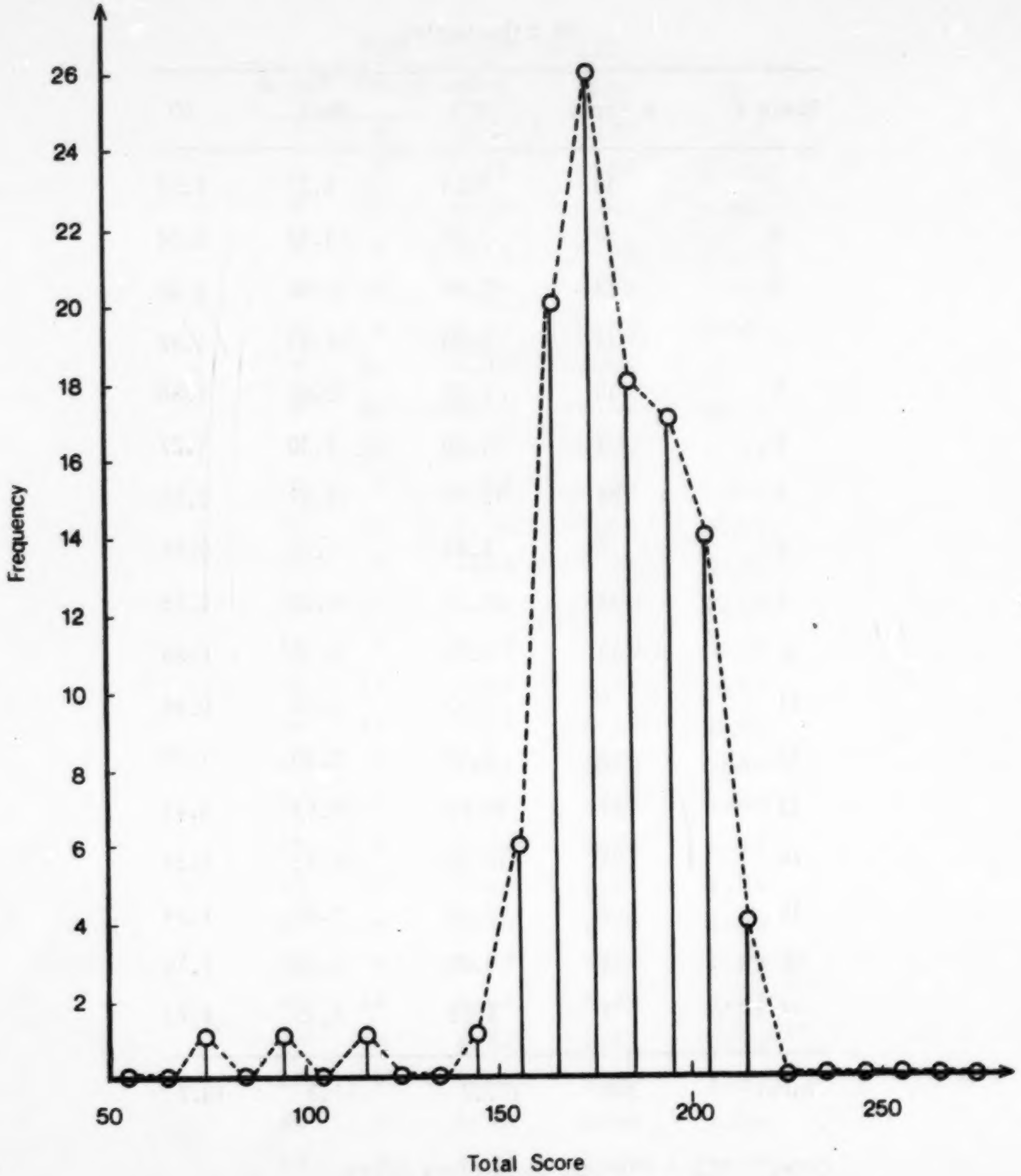
- C. Acquire knowledge and skills essential to incorporating into practice proven new evaluation/treatment modalities
- D. Maintain an on-going library of appropriate journals and books
- E. Maintain membership in professional organizations
- F. Obtain/maintain certification as a PA
- G. Critically review the current literature

XI. PROMOTE ACCEPTANCE OF THE ROLE

- A. Explain role by actions and words to others
- B. Display sensitivity to the partial overlapping and possible sharing of responsibilities with other health professionals
- C. Use formal and informal conflict resolution techniques including adjusting activities, fostering improved working relationships, helping behavior
- D. Transmit reference materials to relevant professionals concerning physician assistant functions and utilization
- E. Assess within the work group the behavior of individuals and group actions to facilitate problem solving or prevent problems from arising
- F. Know and implement strategies useful in gaining acceptance of the role within the community
- G. Give talks to groups interested in the PA concept
- H. Seek out or counsel prospective PA students
- I. Write articles for local newspaper about the PA concept
- J. Submit an article for publication
- K. Initiate contact with other physicians in the area to promote the PA concept
- L. Participate in community health programs
- M. Initiate change in routine protocol

EXHIBIT C

FREQUENCY DISTRIBUTION OF TOTAL SCORES



Note: Sample size = 108
Mean score = 191.3
Minimum competency score = 171.2

EXHIBIT D

MEAN, STANDARD DEVIATION, AND MINIMUM COMPETENCY SCORE FOR THE 17 ROLE SCALES

PA Role Scales

Scale #	# Items	MCS	Mean	SD
1	10	5.34	5.27	1.59
2	16	7.71	11.49	2.24
3	22	12.51	12.44	2.88
4	17	9.84	10.59	1.52
5	15	8.92	10.60	1.88
6	7	3.80	4.30	1.22
7	28	13.75	15.81	3.54
8	5	2.44	1.72	0.94
9	31	17.12	18.24	2.75
10	12	6.28	6.19	1.85
11	5	3.53	3.79	0.96
12	13	6.50	8.36	1.79
13	31	18.45	20.62	3.43
14	24	12.88	14.53	2.34
15	9	5.10	5.65	1.54
16	6	3.20	4.08	1.15
17	7	4.23	4.15	1.23
Total	308	171.2	191.3	18.7

Codes: MCS = Minimum Competency Score
SD = Standard Deviation

EXHIBIT E

MEAN, STANDARD DEVIATION, AND MINIMUM COMPETENCY SCORE FOR THE 28 BODY SYSTEM SCALES

BODY SYSTEM SCALES

Scale#	#Items	MCS	Mean	SD
1	8	4.49	4.86	1.49
2	7	4.13	5.55	1.21
3	10	6.48	6.62	1.45
4	6	4.24	4.48	1.15
5	15	8.54	9.06	2.08
6	9	5.19	5.07	1.50
7	16	8.71	8.89	2.43
8	23	13.03	15.44	2.17
9	18	8.40	10.05	1.85
10	7	3.83	3.93	1.40
11	12	7.76	7.64	1.80
12	9	4.68	5.55	1.38
13	12	6.34	7.57	1.83
14	8	4.31	3.56	1.26
15	12	6.38	6.61	1.97
16	11	5.23	7.13	1.65
17	9	5.06	5.43	1.44
18	6	2.83	3.42	1.40
19	6	2.72	2.82	1.02
20	13	7.22	7.61	1.76
21	6	4.20	4.75	1.06
22	6	3.97	4.18	1.25
23	16	8.62	8.79	2.22
24	3	1.32	2.21	0.73
25	7	3.08	3.40	1.22
26	5	2.93	2.40	1.28
27	5	3.00	3.04	1.24
28	8	3.39	4.06	1.47
TEST TOTAL:	308	171.2	191.3	18.7

Codes: MCS = Minimum Competency Score
SD = Standard Deviation

EXHIBIT F

DISTRIBUTIONS OF ITEM DIFFICULTY AND DISCRIMINATION INDICES FOR TOTAL TEST (SECOND TRY-OUT)

Item Difficulty Distribution (Second Try-Out)

Range	Number of Items	Percent
0.81 to 1.00	34	11
0.61 to 0.80	55	17
0.41 to 0.60	66	21
0.21 to 0.40	86	27
0.00 to 0.20	74	23

Item Discrimination Distribution (Second Try-Out)

Range	Number of Items	Percent
0.81 to 1.00	0	0
0.61 to 0.80	1	0
0.41 to 0.60	29	9
0.21 to 0.40	103	33
0.00 to 0.20	172	55
Below 0.00	10	3

EXHIBIT G

RELIABILITY, STANDARD ERROR OF MEASUREMENT (SEM), AND COEFFICIENT OF ACCEPTANCE FOR THE 17 PA ROLE SCALES

Scale #	# Items	Reliability	SEM	Coefficient Acceptance*
1	10	0.19	1.43	0.58
2	16	0.48	1.62	0.97
3	22	0.51	2.00	0.63
4	17	0.01	1.51	0.75
5	15	0.23	1.64	0.91
6	7	0.16	1.11	0.84
7	28	0.57	2.31	0.80
8	5	0.23	0.83	0.52
9	31	0.32	2.27	0.63
10	12	0.33	1.51	0.58
11	5	0.31	0.80	0.82
12	13	0.30	1.50	0.89
13	31	0.56	2.27	0.78
14	24	0.19	2.11	0.81
15	9	0.37	1.22	0.68
16	6	0.23	1.01	0.84
17	7	0.19	1.11	0.58
Total Test	308	0.90	7.41	1.00

*Note: The Coefficient of Acceptance is the probability of the consistency in classifying an individual as competent or imcompetent on the basis of this test score. It is dependent on the minimum competency score and is usually inflated by consistencies that occur by chance.

EXHIBIT H

RELIABILITY STANDARD ERROR OF MEASUREMENT (SEM), AND COEFFICIENT OF ACCEPTANCE FOR THE 28 BODY SYSTEM SCALES

Scale #	# Items	Reliability	SEM	Coefficient Acceptance*
1	8	0.36	1.19	0.55
2	7	0.31	1.01	0.53
3	10	0.22	1.28	0.76
4	6	0.25	1.00	0.56
5	15	0.33	1.70	0.86
6	9	0.24	1.31	0.53
7	18	0.49	1.74	0.68
8	23	0.21	1.93	1.00
9	18	0.16	1.70	0.86
10	7	0.23	1.23	0.58
11	12	0.32	1.48	0.82
12	9	0.20	1.23	0.53
13	12	0.37	1.45	0.55
14	8	0.14	1.17	0.74
15	12	0.34	1.60	0.66
16	11	0.26	1.42	0.57
17	9	0.25	1.24	0.55
18	6	0.48	1.01	0.90
19	6	0.08	0.98	0.91
20	13	0.19	1.59	0.89
21	8	-0.06	1.09	0.50
22	6	0.39	0.97	0.65
23	16	0.45	1.64	0.73
24	3	-0.01	0.74	1.00
25	7	0.05	1.19	0.72
26	5	0.39	1.00	0.71
27	5	0.52	0.86	0.78
28	8	0.36	1.17	1.62
Total Test	308	0.90	7.41	1.00

*Note: The Coefficient of Acceptance is the probability of the consistency in classifying an individual as competent or incompetent on the basis of this test score. It is dependent on the minimum competency score and is usually inflated by consistencies that occur by chance.

EXHIBIT I

ASSIGNMENT OF SECOND TRY-OUT TEST ITEMS TO TEST SPECIFICATIONS MATRIX

Role Area		1 Knowledge	2 Problem Solving	3 Interpersonal	Total
1	Physician Interdep.	1 (3)	1 (3)	1 (1)	3 (7)
2	Prof. Behavior	(1)	4 (1)	5 (0)	9 (2)
3	Preventive Care	4 (13)	1 (14)	(4)	5 (31)
4	Est. Data Base	46 (19)	29 (20)	(6)	75 (45)
5	Analyze Data	37 (20)	68 (20)	1 (6)	106 (46)
6	Form Health Plan	21 (20)	35 (20)	3 (6)	59 (46)
7	Implement Plan	12 (20)	15 (20)	1 (6)	28 (46)
8	Monitor Plan	9 (6)	11 (20)	1 (6)	21 (45)
9	Est. Interpersonal	(7)	4 (7)	3 (2)	7 (16)
10	Maintain Comp.	(3)	(3)	(1)	(7)
11	Promote PA Role	(4)	1 (4)	1 (1)	2 (9)
Total		130 (129)	169 (132)	16 (33)	315 (300)

Note: Numbers within brackets indicate weights recommended by Test Specifications Committee

EXHIBIT J

AMERICAN ACADEMY OF PHYSICIAN ASSISTANTS
SELF ASSESSMENT EXAMINATION -- FORM A

```
THIS REPORT IS PREPARED FOR AAPA ID #
WHO COMPLETED THIS EXAMINATION ON    7/13/79.


YOUR TOTAL RAW SCORE IS 165  OUT OF 308; YOU OMITTED    2 ITEMS.
MINIMUM COMPETENCY TOTAL SCORE IS 171.2.
MEAN SCORE OF NATIONAL SAMPLE OF PAS IS 187.8, WITH STD DEV= 23.7

PRTNTED BELOW IS YOUR RESPONSE TO EACH ITEM.
THE CORRECT RESPONSE IS SHOWN WITHIN BRACKETS ONLY WHEN DIFFERENT.
ASTERISK INDICATES ITEM IS NOW DELETED FROM EXAM.


SECTION 1

  1 A         2 A         3 E         4 B(E)      5 B(D)      6 E         7 D(E)      8 B(D)
  9 D(E)     10 C(E)     11 A(C)     12 B        13 B        14 A        15 E        16 B(E)
 17 E        18 B        19 B        20 A(E)     21 C        22 A        23 B(C)     24 A(E)
 25 D(B)     26 E *      27 B(A)     28 D(C)     29 B        30 B        31 A        32 E
 33 A(E)     34 E(B)     35 A        36 D(E)     37 B(A)     38 B(E)     39 A        40 C(A)
 41 E        42 C(B)     43 E(B)     44 A(E)     45 D(E)     46 B        47 C        48 A
 49 C(A)     50 C(B)     51 A        52 A(D)     53 A        54 E(D)     55 A        56 B
 57 A        58 B        59 D        60 A(C)     61 D(A)     62 C(E)     63 D(C)     64 C(B)
 65 C(B)     66 D        67 B *      68 E(D)     69 A *      70 B(D)     71 D        72 B(A)
 73 D(E)     74   (A)    75 E(D)     76 A        77 A        78 C(E)     79 C(D)     80 D(B)
 81 D        82 D(E)     83 D        84 E        85 A(C)     86 C        87 E *      88 C
 89 C        90 D        91 C(D)     92 B        93 E(B)     94 C(A)     95 E(D)     96 E(A)
 97 B        98 A(B)     99   (D)   100 A       101 A       102 D       103 C       104 D
105 A(E)    106 C(E)    107 C       106 D       109 E(B)    110 B       111 D       112 E
113 D       114 B(A)    115 C(A)    116 A       117 B       118 B       119 C(D)    120 D(B)
121 B(E)    122 D       123 D       124 D       125 D       126 A       127 B(A)    128 D(A)
129 B(A)    130 D       131 E(D)    132 C       133 D       134 D       135 D(C)    136 D
137 A(E)    138 C(B)    139 C       140 A(C)    141 A       142 E       143 D       144 D
145 E(A)    146 E       147 D       148 E       149 B(C)    150 C       151 D       152 C
153 A(E)    154 B(D)    155 B       156 B       157 D       156 C       159 A       160 E

SECTION 2

  1 B         2 D         3 E(B)      4 D         5 D(B)      6 B(C)      7 D         8 D
  9 B *      10 C        11 D        12 A        13 D(A)     14 B        15 C        16 C
 17 C        18 C(E)     19 A        20 B(C)     21 D(C)     22 B(A)     23 A(D)     24 A(E)
 25 A(B)     26 D        27 D(B)     28 D        29 B(A)     30 C        31 C        32 A
 33 A(D)     34 B(D)     35 B        36 A        37 C        38 D        39 C(E)     40 D
 41 D        42 C        43 E        44 A(E)     45 C(B)     46 A        47 E        48 D(A)
 49 D(C)     50 A(C)     51 B(E)     52 E(A)     53 B(E)     54 D(E)     55 E(C)     56 B(C)
 57 E        58 D        59 E        60 D(E)     61 B        62 D        63 C        64 A
 65 A(B)     66 B        67 A        68 D        69 E        70 A(C)     71 B        72 B(D)
 73 D        74 C(D)     75 D        76 A        77 C(B)     78 C(E)     79 C        80 D
 81 D *      82 D        83 C(B)     84 C(D)     85 C(B)     86 B(D)     87 A        88 D(E)
 89 B        90 B(A)     91 B(E)     92 A(B)     93 C(B)     94 A        95 D        96 C
 97 C(B)     98 D        99 A       100 B       101 D       102 B(E)    103 A       104 E(A)
105 E(D)    106 B(C)    107 C       108 D *     109 C(D)    110 C *     111 C(B)    112 C(D)
113 B       114 A(B)    115 D       116 B       117 A       118 B(C)    119 C       120 C
121 C       122 C       123 D       124 D       125 B       126 E(B)    127 B       128 B(C)
129 C       130 A(B)    131 A(C)    132 A(D)    133 D(C)    134 C       135 A(D)    136 C(E)
137 A(D)    138 A(B)    139 A(E)    140 D       141 B       142 C       143 E(D)    144 B
145 B(C)    146 C       147 A       148 D(E)    149 D(E)    150 A       151 C       152 E
153 A(B)    154 A       155 A(D)

                                                                                   CONTINUED
```

EXHIBIT J CONTINUED

PAGE 2 OF REPORT FOR AAPA ID NUMBER

BROKEN DOWN BY ROLE AREA AND BODY SYSTEM,
YOUR EXAMINATION SCORES WOULD APPEAR AS FOLLOWS:

SCALE	TOTAL ITEMS	YOUR SCORE	MIN COMP
1 PROFESSIONAL ROLE	10	6	5.3
2 INTERPERSONAL BEHAV.	16	10	7.7
3 GATHER DATA--RESP,CV	22	8	12.5
4 GATHER DATA--GI,GU	17	10	9.8
5 GATHER DATA--PSY,NEU	15	10	8.9
6 ANALYZE DATA--ENDO	7	5	3.8
7 ANALYZE DATA-RESP,CV	28	9	13.7
8 ANALYZE DATA--HEMA	5	2	2.4
9 ANALYZE DATA--GI,GU	31	17	17.1
10 ANALYZE DATA-PSY,NEU	12	4	6.3
11 FORM PLAN--MUSC-SKEL	5	3	3.5
12 HANDLE DERM PROBLEM	13	6	6.5
13 MANAGE PTS--RESP,CV,	31	18	18.4
14 MANAGE PTS--GI,GU	24	13	12.9
15 MANAGE PTS--REPRODUC	9	7	5.1
16 MANAGE PTS--PSY,NEUR	6	5	3.2
17 HANDLE PHARM PROBLEM	7	4	4.2

CONTINUED

EXHIBIT J CONTINUED

PAGE 3 OF REPORT FOR AAPA ID NUMBER

BROKEN DOWN BY BODY SYSTEM AND MEDICAL INTERVENTION, YOUR EXAMINATION SCORES WOULD APPEAR AS FOLLOWS:

SCALE	TOTAL ITEMS	YOUR SCORE	MIN COMP
EMERGENCY			
1 CARDIO-VASCULAR	8	4	4.5
2 RESPIRATORY	7	4	4.1
3 GASTRO-INTESTINAL	10	6	6.5
4 NEUROLOGY	6	5	4.2
ACUTE			
5 CARDIO-VASCULAR	15	6	8.5
6 MUSCULO-SKELETAL	9	4	5.2
7 RESPIRATORY	16	7	8.7
8 GASTRO-INTESTINAL	23	15	13.0
9 GENITO-URINARY	18	11	8.4
10 NEUROLOGY	7	3	3.8
11 PHARMACOLOGY	12	8	7.8
12 DERMATOLOGY	9	5	4.7
13 EYES & ENT	12	5	6.3
14 HEMATOLOGY	8	3	4.5
15 REPRODUCTIVE	12	7	6.6
16 PSYCHO-SOCIAL	11	7	5.2
CHRONIC			
17 CARDIO-VASCULAR	9	4	5.1
18 MUSCULO-SKELETAL	6	4	2.8
19 RESPIRATORY	5	3	2.7
20 GASTRO-INTESTINAL	13	4	7.2
21 GENITO-URINARY	8	5	4.2
22 NEUROLOGY	6	4	4.0
23 ENDOCRINE	16	8	8.6
24 DERMATOLOGY	3	1	1.3
25 EYES & ENT	7	2	3.1
26 HEMATOLOGY	5	2	2.9
27 REPRODUCTIVE	5	4	3.0
28 PSYCHO-SOCIAL	8	3	3.4

CONTINUED

EXHIBIT J CONTINUED

```
PAGE 4 OF REPORT FOR AAPA ID NUMBER

COMPARED WITH YOUR PRACTICE (P) PROFILE (TYPES OF PATIENTS SEEN)
    AND WITH YOUR CME NEEDS (N) PROFILE (AREAS SOUGHT TO LEARN),
            YOUR EXAM (E) SCORES WOULD SHOW UP AS FOLLOWS:

EMERGENCY
1 CARDIO-VASCULAR    P----------            3 GASTRO-INTESTINAL P--------
                     N........                                  N........
                     E******                                    E****

2 RESPIRATORY        P----------            4 NEUROLOGY         P----------
                     N..........                                N........
                     E****                                      E******

ACUTE                                       CHRONIC
5 CARDIO-VASCULAR    P------------          17 CARDIO-VASCULAR  P----------
                     N..........                                N..........
                     E****                                      E******

6 MUSCULO-SKELETAL   P----------            18 MUSCULO-SKELETAL P--------
                     N..........                                N..........
                     E******                                    E******

7 RESPIRATORY        P----------            19 RESPIRATORY      P----------
                     N..........                                N..........
                     E******                                    E******

8 GASTRO-INTESTINAL  P--------              20 GASTRO-INTESTINAL P----------
                     N..........                                N............
                     E**                                        E****

9 GENITO-URINARY     P----------            21 GENITO-URINARY   P----------
                     N..........                                N..........
                     E****                                      E****

10 NEUROLOGY         P----------            22 NEUROLOGY        P----------
                     N............                              N............
                     E******                                    E******

11 PHARMACOLOGY      P----------            23 ENDOCRINE        P----------
                     N............                              N..........
                     E****                                      E******

12 DERMATOLOGY       P--------              24 DERMATOLOGY      P--------
                     N..........                                N..........
                     E****                                      E******

13 EYES & ENT        P--------              25 EYES & ENT       P----------
                     N..........                                N..........
                     E****                                      E******

14 HEMATOLOGY        P----------            26 HEMATOLOGY       P----------
                     N............                              N............
                     E******                                    E********

15 REPRODUCTIVE      P----------            27 REPRODUCTIVE     P----------
                     N..........                                N..........
                     E******                                    E******

16 PSYCHO-SOCIAL     P----------            28 PSYCHO-SOCIAL    P--------
                     N............                              N..........
                     E****                                      E******
```

EXHIBIT K

INTERPRETIVE LEAFLET

Competency-Based Self-Assessment for Physician Assistants
developed by the
American Academy of Physician Assistants (AAPA)
Form A, 1979

HOW TO INTERPRET YOUR REPORT

(an interpretive leaflet for the physician assistant)

Introduction

The self-assessment examination which you recently completed and the enclosed computer-generated report were developed by the AAPA under Contract No. HRA 231-76-0053 with the guidance of Ayres D'Costa, Ph.D., Associate Professor, Health Professions Education, The Ohio State University, Columbus, Ohio. Major support for the development of the exam was provided by an Item Writers Committee and a Test Specifications Committee whose members are PA practitioners and PA educators.

This *Interpretive Leaflet* is designed to assist you in deriving some benefits from the exam. Comparisons between your scores as an individual and the scores of PAs as a group are not emphasized, because we want the self-assessment exam to be criterion-referenced. Our criterion is PA competence, and we would like to help you work towards this goal.

In order to assure that the exam does cover the areas of competence expected of PAs, we utilized the Role Delineation for the Physician Assistant recently developed by the AAPA. Eleven major responsibilities, or role areas, span the realm of tasks which a PA should be competent to perform. The items included in the self-assessment exam were considered to be the best items available to AAPA for representing the critical PA behaviors relative to competent performance in the 11 role areas. (For a complete discussion of the item generation and test development process, see AAPA's Final Report to HRA on this project, *Volume III: Development of a Self-Assessment Examination for Physician Assistants* .)

Certain limitations of the exam must be pointed out in interpreting your exam results. The exam was under development when you completed it and is, therefore, subject to further review and refinement. Moreover, we have not yet analyzed the extent to which the sample who actually completed the exam is representative of the PA population. As the self-assessment program gains in experience, the quality and quantity of the test items available to us will improve, and the results you will receive in the future will carry greater credibility.

Technical Information

The technical information included in this leaflet is not entirely necessary in order to interpret your exam results. Such information is provided within brackets [ ] for those readers who might be interested in it.

[The reliability of the total test is 0.90, with a standard error of 7.41. These figures are generally considered respectable by measurement professionals. However, the attached computer-generated report also includes scores on scales (described below) derived from subsets of items from the total test. The reliabilities of the scales can therefore be expected to be lower than the reliability of the total test. The reliabilities of these scales are presented in Tables 1 and 2 so that you can use the necessary caution in using this data in your self-assessment. In a criterion-referenced exam the coefficient of agreement serves as the more appropriate index of reliability. See Tables 1 and 2.]

How to Interpret the Report

The computer-generated report consists of four pages. Among the information reported are: your response and the correct response to each item, your scores on scales derived from subsets of items, and a graphic comparison of your practice characteristics, your CME needs, and your exam scores. Detailed instructions on how to interpret each page of the report are given below.

How to Interpret Page 1 . Your *Total Raw Score* is the total number of test questions you answered correctly. You may recall that there were 315 questions on the test. However, as a result of item analyses and further review, seven items were discarded from the self-assessment exam and were not scored or included in any examination statistics. Therefore, the highest score one could attain is 308.

In order to judge the adequacy of your total raw score, you should compare it to the *Minimum Competency Score* . This score is based upon the expert judgements of some of the PAs and PA educators who developed the exam. The word "minimum" is intended to convey a basic level of competence for an entry level generalist PA. Following the minimum competency score is the number of items you *Omitted* on your answer sheet.

EXHIBIT K CONTINUED

Table 1

RELIABILITY, STANDARD ERROR OF MEASUREMENT (SEM), AND COEFFICIENT OF ACCEPTANCE FOR THE 17 PA ROLE SCALES

Scale #	# Items	Reliability	SEM	Coefficient Acceptance*
1	10	0.19	1.43	0.58
2	16	0.48	1.62	0.97
3	22	0.51	2.00	0.63
4	17	0.01	1.51	0.75
5	15	0.23	1.64	0.91
6	7	0.16	1.11	0.84
7	28	0.57	2.31	0.80
8	5	0.23	0.83	0.52
9	31	0.32	2.27	0.63
10	12	0.33	1.51	0.58
11	5	0.31	0.80	0.82
12	13	0.30	1.50	0.89
13	31	0.56	2.27	0.78
14	24	0.19	2.11	0.81
15	9	0.37	1.22	0.68
16	6	0.23	1.01	0.84
17	7	0.19	1.11	0.58
Total Test	308	0.90	7.41	1.00

*Note: The Coefficient of Acceptance is the probability of the consistency in classifying an individual as competent or incompetent on the basis of this test score. It is dependent on the minimum competency score and is usually inflated by consistencies that occur by chance.

Table 2

RELIABILITY, STANDARD ERROR OF MEASUREMENT (SEM), AND COEFFICIENT OF ACCEPTANCE FOR THE 28 BODY SYSTEM SCALES

Scale #	# Items	Reliability	SEM	Coefficient Acceptance*
1	8	0.36	1.19	0.55
2	7	0.31	1.01	0.53
3	10	0.22	1.28	0.76
4	6	0.25	1.00	0.56
5	15	0.33	1.70	0.86
6	9	0.24	1.31	0.53
7	18	0.49	1.74	0.68
8	23	0.21	1.93	1.00
9	18	0.16	1.70	0.86
10	7	0.23	1.23	0.58
11	12	0.32	1.48	0.82
12	9	0.20	1.23	0.53
13	12	0.37	1.45	0.55
14	8	0.14	1.17	0.74
15	12	0.34	1.60	0.66
16	11	0.26	1.42	0.57
17	9	0.25	1.24	0.55
18	6	0.48	1.01	0.90
19	6	0.08	0.98	0.91
20	13	0.19	1.59	0.89
21	8	-0.06	1.09	0.50
22	6	0.39	0.97	0.65
23	16	0.45	1.64	0.73
24	3	-0.01	0.74	1.00
25	7	0.05	1.19	0.72
26	5	0.39	1.00	0.71
27	5	0.52	0.86	0.78
28	8	0.36	1.17	1.62
Total Test	308	0.90	7.41	1.00

EXHIBIT K CONTINUED

The *Mean Score* is commonly called the arithmetic average; it indicates how the national sample of PAs scored on the exam. You should use the standard deviation (*Std Dev*) along with the mean to appreciate how far the scores of the entire sample are spread out from the mean score (i.e., are distributed).

The extensive printouts under *Section 1* and *Section 2* refer to the test items. Your response is printed after each item number. A letter in parentheses is the correct response and appears only if you answered the item incorrectly. The items you did not answer are left blank. The items that have been discarded from the exam are marked with an asterisk.

Note the items that you answered incorrectly. You might wish to go back to the test questions. (If you no longer have your test booklets, write the AAPA for another set.) Do you agree with the correct answer? Maybe you don't for a very good reason. We would like to hear your reasons and your comments. Remember that our item writers and committee members are also PAs like you. Maybe we goofed! Perhaps you agree with our experts and would like to undertake some self-learning. We have not been able to develop such learning packages yet, but would like to. Hearing of your interest will help us. Do write and let us know your specific needs.

How to Interpret Page 2 . The 17 scales on this page were derived from the Role Delineation for the Physician Assistant. There are 11 role areas and several of the areas are subdivided by 13 body systems. At this point in the development of items for a self-assessment exam, not enough items are available to form a scale for each of the role areas and body systems. Therefore, some of the scales on this page represent more general areas of competence derived by a meaningful combination of the 11 role areas. For some role areas, that are well represented on the self-assessment exam, it was possible to derive more than one scale so as to represent competence within various body systems. At this stage, there is a limited number of role area - by - body system scales available. The 11 role areas, the 13 body systems, and a graphic representation of these 17 scales are presented in Table 3. You will note that one of the 11 role areas on the role delineation -- promote preventive health care -- does not have any scale on the exam, as not enough test items have been generated for this area.

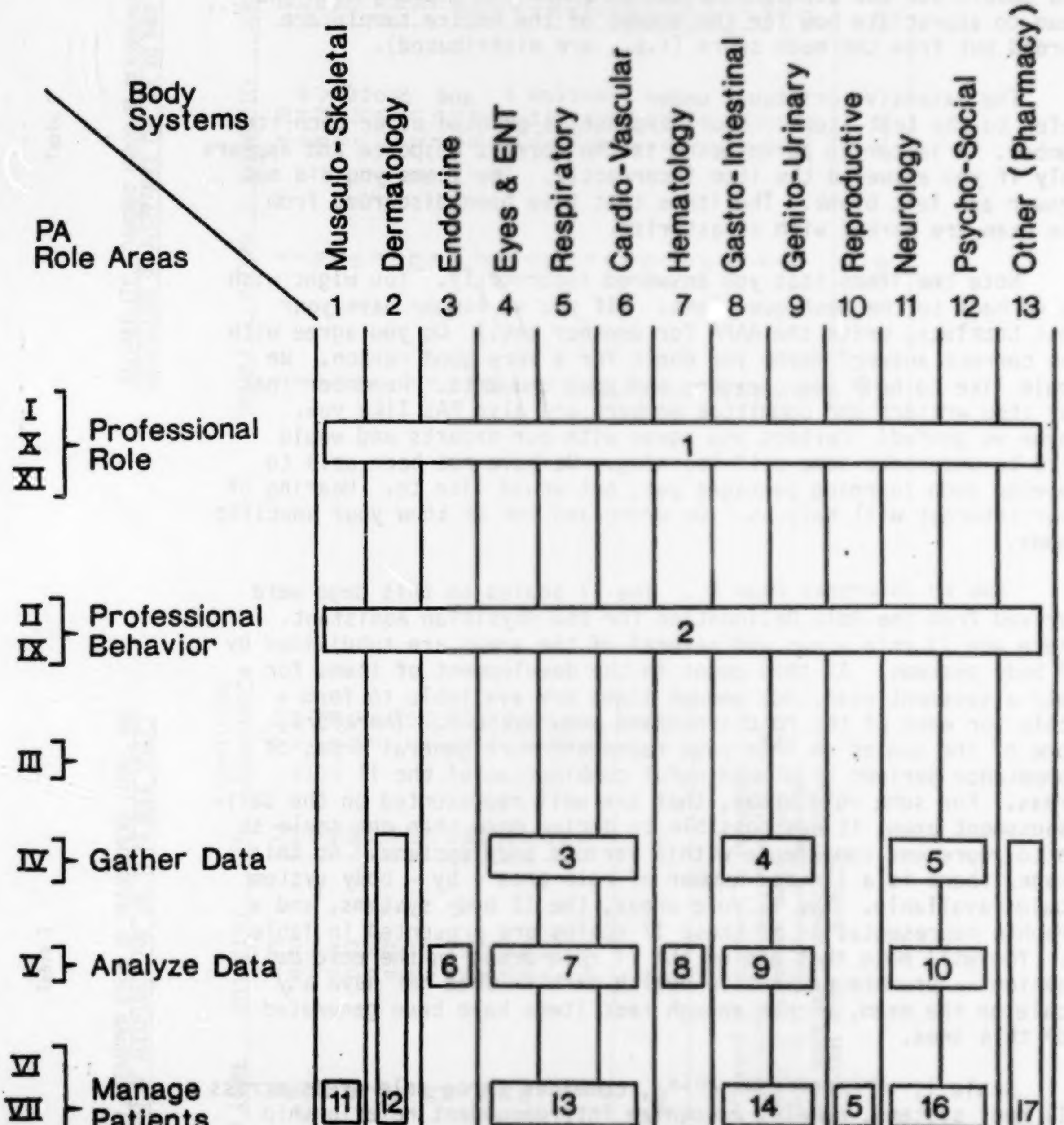
Scale 1, *Professional Role* , combines three role areas across all body systems, namely, recognize interdependent relationship with supervising physician, maintain competency, and promote acceptance of the role.

Scale 2, *Interpersonal Behavior* , combines two role areas across all body systems, namely, demonstrate professional behavior

EXHIBIT K CONTINUED

Table 3

MAP OF THE 17 ROLE AREAS X BODY SYSTEM SCALES



Body Systems
PA Role Areas
Musculo-Skeletal
Dermatology
Endocrine
Eyes & ENT
Respiratory
Cardio-Vascular
Hematology
Gastro-Intestinal
Genito-Urinary
Reproductive
Neurology
Psycho-Social
Other (Pharmacy)
1
2
3
4
5
6
7
8
9
10
11
12
13
I
X
XI
Professional Role
1
II
IX
Professional Behavior
2
III
IV
Gather Data
3
4
5
V
Analyze Data
6
7
8
9
10
VI
VII
VIII
Manage Patients
11
12
13
14
5
16
17

EXHIBIT K CONTINUED

and establish effective interpersonal relationships with patients, professionals, and others.

Scales 3, 4, and 5, *Gather Data* , relate to competence in one role area (establish health status data base) separately for three body systems.

Scales 6 through 10, *Analyze Data* , relate to competence in one role area (analyze the health status data base) for various body systems.

Scale 11 relates to one role area (*formulate health management plan*) for one body system (*musculoskeletal*).

Scale 12 combines five role areas (establish and analyze data base and formulate, implement, and monitor health management plan) relevant to *dermatology* problems.

Scales 13 through 16 (*Manage Patients*) combine three role areas (formulate, implement, and monitor health management plan) for various body systems.

Scale 17 combines the five role areas of Scale 12 relative to *pharmacological* problems.

Three scores are provided for each of these 17 scales. The first indicates the *total* number of *items* included in the scale and is the maximum possible score on the scale. The second is *Your Score* and it indicatthe number of these items that you answered correctly. The third is the *Minimum Competency* score which is a tentative benchmark for comparison with your scale score. [Do not forget possible problems of reliability of scales with a small number of items. The reliability index and the standard error for each scale are listed in Table 1. A reliability of 0.70 is the minimum we would like to see. However, several factors, such as number of items and group diversity, affect the reliability index. For this reason, it is difficult to provide a general rule for interpreting this index.]

How to Interpret Page 3 . The 28 scales on this page were derived from available items representing body systems as well as type of medical intervention. These scales provide another way of interpreting your exam results. The 13 categories of body system, the four types of intervention, and a graphic representation of the 28 scales are presented in Table 4. You will see from the Table that "Acute" and "Maintenance" are combined, as there are not yet sufficient exam items to represent both types of intervention. Also, not all body systems are represented relative to "Emergency" interventions. Both "Acute" and "Chronic" appear to be well represented. Scale 11 does not represent a body system. However, sufficient exam items were found relevant to drug information (dosage, side effects, etc.) to include a scale for "Acute Pharmacology."

EXHIBIT K CONTINUED

Table 4

MAP OF THE 28 BODY SYSTEMS x MEDICAL INTERVENTION SCALES

	Emergency	Acute & Maintenance	Chronic
1. Musculo-Skeletal		6	18
2. Dermatology		12	24
3. Endocrine			23
4. Eyes & ENT		13	25
5. Respiratory	2	7	19
6. Cardio-Vascular	1	5	17
7. Hematology		14	26
8. Gastro-Intestinal	3	8	20
9. Genito-Urinary		9	21
10. Reproductive		15	27
11. Neurology	4	10	22
12. Psycho-Social		16	28
13. Other (Pharmacy)		11	

EXHIBIT K CONTINUED

The format of information on page 3 is identical to that of Page 2. Again three scores are provided. Similar cautions are urged in utilizing this information in self-assessment. The reliability index and standard error for each scale are listed in Table 2.

How to Interpret Page 4 . We expect that you will find this page very interesting and useful to you in your self-assessment. Instead of actual raw scores, page 4 presents a graphic representation of your scores on the 28 scales from page 3. Page 4 is arranged to allow you to make comparisons among your "practice characteristics (P)," your "CME needs (N)," and your "exam scores (E)." The P and N representations were derived from your responses to Section 3 of the exam, where questions were of the type "How many patients do you see for certain types of care?" (for practice characteristics) and "How comfortable do you feel about your performance with patients requiring certain types of care?" (for CME needs).

We believe that page 4 will be helpful to you in planning your CME activities. You can make two kinds of visual comparisons on page 4. You can compare your practice, your needs, and your exam score on the same scale. Also, you can compare your practice across all the scales to find out which types of patient problems you see relatively more and less often in your practice. Similarly, you can compare your CME needs and exam scores across all scales.

[Your scale scores on page 3 and their graphic representations on page 4 cannot be compared directly, as the latter represent standard scores. Likewise, the practice characteristics scores and the CME needs scores are standard scores. All of these standard scores have a mean of 5.0, a standard deviation of 0.5, and a range 1 to 9. The standard scores are based upon the national sample of PAs who completed the exam in June 1979.]

Here are some suggestions for utilizing the information on this page:

i) Identify your high need areas. Do these needs relate to your practice characteristics?

ii) Compare your high need areas to your exam scores. Are there some mismatches? In which direction -- are there some high needs where exam scores are low; or perhaps low needs scores where exam scores are high? Perhaps your needs are high even when your exam scores are high. Is this indicative of your high goals and ambition in this area?

iii) Are your need scores consistently high? Try to understand why. Motivation? Pressure?

EXHIBIT K CONTINUED

Conclusion

This interpretive leaflet has merely indicated some approaches to using the AAPA Self-Assessment Report. You will undoubtedly wrestle with additional questions and come up with other and more imaginative ways for using the information in this Report. Once again, we urge caution because this program is very much in an experimental stage.

Finally, the AAPA congratulates you for taking the self-assessment exam and hopes that you will find this Report useful.

best copy available

GPO 872 376

copy covers+T

BHPr

U.S. DEPARTMENT OF
HEALTH AND HUMAN SERVICES
Public Health Service
Health Resources Administration
Bureau of Health Professions

DHHS Publication No. (HRA) 80-38